Implementing the Environmental Sustainability Gap framework to monitor the environmental sustainability of nations

Arkaitz Usubiaga Liaño

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of
University College London

Institute for Sustainable Resources
Bartlett Faculty of the Built Environment
University College London

January 14, 2022

THE BARTLETT
SCHOOL OF ENVIRONMENT,
ENERGY AND RESOURCES

UNIVERSITY COLLEGE LONDON

Arkaitz Usubiaga Liaño

Supervisors:

Prof. Paul Ekins
UCL Institute for Sustainable Resources

Prof. Raimund Bleischwitz
UCL Institute for Sustainable Resources

Declaration

I, Arkaitz Usubiaga Liaño, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgments

First and foremost, I would like to thank my primary supervisor, Paul Ekins, for his continued support during all these years. Throughout the PhD, Paul has also been an inspiration both at an academic and at a personal level. He was one of the main reasons for me to join ISR and he turned out to the best supervisor I could have ever expected. Honestly, I cannot put into words how grateful I am for everything you have done. Paul, I hope this thesis meets the expectations you had when you formulated the original SGAP approach.

I would also like to thank my secondary supervisor, Raimund Bleischwitz, for his advice and for having the door always open for me. Thanks as well to Oskar Lecuyer (AFD) for pursuing, promoting and funding additional ESGAP work, to Alison Fairbrass (ISR) for working side by side during these last years, to Jock Martin (EEA) and his unit for facilitating discussions around the ESGAP framework, to Thomas Scheusnar (UBA), Panos Panagos (JRC) and Nihat Zal (EEA) for sending additional data that was not available online at the time, to the late Georgina Mace (CBER) for her kindness and exchanges in early stages of the PhD, and lastly to my examiners Tomáš Hák and Dan Obsorn for reading the thesis and providing valuable feedback.

I am eternally grateful to those who made my life in London much more enjoyable. Working in an institute with such academically and culturally diverse researchers has been very enriching. I have had the chance to work and teach alongside a very committed and skilled group of people, which has been a fantastic experience. At ISR, my fellow PhD students – in particular Wan-Ting, Tony, Theo, Alex, Victor, Stijn and Ken – made the lunch breaks something to look forward to and made Friday afternoons at the pub so much fun. At home, the long conversations with Monique about the PhD life and many unexpected topics really helped to take the load off the working routine. Fuera del trabajo, he disfrutado cada momento con mis amigos de Sectarismo en Londres. Fue una suerte increíble el haberos encontrado allí.

También quería agradecer a todos mis amigos de Bilbao por hacerme sentir que no había distancia entre nosotros. En especial, a la cuadrilla Ramón Peña y a los Tomateros, a la gente del equipo de basket y a las amigas de la uni. Por supuesto, quiero dar las gracias a mi familia, y en especial a Ama, Aita y a Ibai. Durante los más de diez años que he vivido fuera, cada visita y cada reunión familiar me daban la fuerza y el cariño necesario para seguir adelante. Eskerrik asko guztioi! Por último, quiero agradecer a Diana el hacerme feliz cada día.

Publications based on this PhD

Peer-reviewed papers

- **Usubiaga-Liano, A.**, Mace, G. M. and Ekins, P. (2019). Limits to agricultural land for retaining acceptable levels of local biodiversity. Nature Sustainability, 2(6), 491-498.
- **Usubiaga-Liaño, A.**, and Ekins, P. (2021). Time for Science-Based National Targets for Environmental Sustainability: An Assessment of Existing Metrics and the ESGAP Framework. Frontiers in Environmental Science, 524.
- **Usubiaga-Liano, A.**, and Ekins, P. (2021). Monitoring the environmental sustainability of countries through the strong environmental sustainability index. Ecological Indicators, 132, 108281.

Book chapters

- Ekins, P. and **Usubiaga, A.** (2019). Brundtland+ 30: the continuing need for an indicator of environmental sustainability. In: Linnerud, K., Holden, E., Langhelle, O., Banister, D. and Meadowcroft, J. (eds.). What Next for Sustainable Development? Edward Elgar Publishing.

Reports

- Ekins, P., Milligan, B. and **Usubiaga-Liaño, A.** (2019). A single indicator of strong sustainability for development: Theoretical basis and practical implementation. AFD Research Papers, No. 2019-112.
- Fairbrass, A., **Usubiaga-Liaño, A.**, Ekins, P. and Milligan, B. (2020). Data opportunities and challenges for calculating a global Strong Environmental Sustainability (SES) index. AFD Research Papers, No. 2020-133.

Abstract

Countries lack resonant metrics to monitor environmental sustainability from a strong sustainability perspective. Building on the Sustainability Gap approach, which was developed in the late 1990s to address this indicator gap, this thesis formulates the Environmental Sustainability Gap (ESGAP) framework with a stronger focus on implementation.

ESGAP comprises two novel indices of environmental sustainability: the Strong Environmental Sustainability Index (SESI) and the Strong Environmental Sustainability Progress Index (SESPI). SESI measures the performance of 21 natural capital indicators against science-based reference values of environmental sustainability that reflect whether the environmental functions provided by natural capital are threatened. Based on observed and desired trends, SESPI describes whether the country is making progress towards, or away environmental sustainability as defined by those environmental sustainability reference values. The analysis focuses on European countries due to good data availability.

European countries perform quite poorly with SESI, which indicates that several environmental functions are threatened. Broadly speaking, European countries perform better in the functions related to the provision of natural resources and human health and welfare, but get lower scores in the functions associated with pollution and life support systems. As shown by SESPI, current trends are also insufficient to reach environmental standards by 2030, although relevant differences emerge depending on the countries and indicators. The results contrast with the generally high performance attributed to European countries in other environmental indices such as the Environmental Performance Index or the Sustainable Development Goals (SDG) Index. A qualitative assessment of the environmental SDG indicators suggests that the SDG indicators fail to represent strong sustainability, which can ultimately lead to misleading messages around environmental sustainability.

Combined, SESI and SESPI can make the messages on environmental sustainability more digestible to relevant audiences, while complementing existing metrics, including those used in the context of the Beyond GDP literature.

Impact statement

The main contribution of this thesis relates to the conceptualisation and quantification of environmental sustainability, which has impacts on research and policy as follows.

The ESGAP framework advances the conceptualisation of environmental sustainability when seen through the lens of strong sustainability. ESGAP makes the concept of environmental sustainability more specific and proposes key criteria that relevant indicators need to fulfil. The definition of environmental sustainability conditions through environmental standards is a central aspect of such indicators. In this vein, the literature review presented in chapter 3 provides an extensive overview of environmental standards – alongside relevant knowledge gaps – that was so far lacking. Their use to define suitable environmental sustainability reference values at the national level is related (although with many caveats) to the use of Planetary Boundaries framework at the global level and the Science-Based Targets Initiative at company level, and therefore should be seen as complementary.

From a practical perspective, SESI and SESPI advance the measurement of environmental sustainability compared to existing metrics. In this context, the choices made during the construction of the indices are closely aligned with the theoretical framework, something often missing in other sustainable development and environmental indices. This approach allows capturing key aspects of environmental sustainability that can be otherwise omitted. Given that the focus in chapters 4 and 5 is set on European countries, the results presented can inform the work of the European Environment Agency or complement their indicator-based assessments. Of special interest would be to compare how environmental targets and environmental standards differ in practice, and to understand how country performance and the perception of success varies depending on whether a policy perspective or a strong sustainability perspective is adopted. This could have relevant implications for target setting in environmental policy making.

The key features of the ESGAP framework and the construction of SESI have been documented in peer-reviewed papers and other reports. This work has been used to guide the implementation of ESGAP in different countries (e.g. New Caledonia, Kenya, Vietnam, China, Japan and the Bahamas) as part of other projects. In these case studies, country-specific versions of SESI have been computed, thereby considering the national context and data capabilities. The policy implications in those countries have been considered as well.

This thesis also elaborates on the limitations of the SDG indicators for monitoring environmental sustainability and progress towards it. This highlights the value added of the ESGAP metrics and how they can complement SDG assessments to incorporate a strong sustainability perspective on the environment. In this line, the ESGAP framework has been featured in a recent report by the United Nations Environment Programme in which progress towards the environmental SDGs was assessed.

Table of contents

List of figures

List of tables

List of boxes

1. Introduction

1.1. Background

1.1.1. Humans and nature

Human well-being has always rested to a higher or lower degree on the benefits obtained through interacting with nature. These benefits include, for instance, basic processes that regulate the Earth System, goods such as food or fresh water that are indispensable for our subsistence, and materials that represent the physical foundations of modern infrastructure. Through the interaction with the natural environment, humans have played a significant role in altering Earth's ecosystem for several millennia. Massive predation of megafauna at around 13.800 B.P. and the use of fire reshaped the vegetation cover across continents (Ellis 2018). Deforestation resulting from early agricultural expansion and methane emissions from inefficient wet rice agriculture at 5.000 B.P. led to observable changes in global temperatures (Smith and Zeder 2013). Since the beginning of the industrial revolution in the mid-1700s, and especially since the great acceleration in the mid-1900s, human pressure on the environment has increased exponentially (Steffen et al. 2015a). It is so well established that humans are currently a 'great force of nature' that it is under consideration to name a new geological epoch – the Anthropocene – after humans (Ellis 2018).

Currently, humanity faces multiple environmental challenges that require appropriate governance structures at all levels, from local to global. There have been stories of success such as the reduction of acid rain in Europe since the 1990s (EEA 2016a) or the ongoing recovery of the ozone layer (Solomon et al. 2016), but there is ample evidence of widespread environmental degradation (Millennium Ecosystem Assessment 2005; IPCC 2013; UN Environment 2019b). The 2019 Global Environmental Outlook report (UN Environment 2019b) compiled the latest evidence on the status of the environment and concluded that current paths of economic development will lead to unprecedented hardship for billions of people, as the most basic systems that support human life on Earth start to unravel. From this outlook it is clear that our current development model is far from being environmentally sustainable. Reversing these trends will not be possible without additional policies that promote a transition to a low-carbon circular economy and radical changes to our consumption patterns (UN Environment 2019b).

1.1.2. The role of environmental metrics in policy

Building bridges between science and policy is critical in this context. The phrase "we cannot manage what we cannot measure" has become part of the vocabulary of those using quantitative tools to produce policy-relevant information. Of course, the statement cannot be taken as an absolute truth. The increasing reliance on quantification in policy can also lead to unintended misuse or even politicisation of data with negative consequences (Radermacher 2019; Umbach 2020). Nonetheless, it is generally agreed that it is relevant to have a clear and scientifically sound information base around which decisions can be made. As Esty (2018, p. 496) argued, "better metrics and data analysis can make the invisible visible, the intangible tangible, and the complex manageable. The 'realization' effect of numbers can be transformative".

Environmental information is commonly structured in a three-level pyramid that comprises data at the bottom, accounting systems in the middle and metrics at the top (Figure 1a). Data represents information compiled by statistical offices or by other sources that can be reused for statistical purposes (Eurostat 2014b). In itself, data refers to independent bits of information with limited coherence. Accounting systems allow overcoming the lack of coherence by providing a set of rules for data compilation and a structure around which data can be organised in a way that makes it more consistent, comparable and usable for analysis (Eurostat 2014b). At the top, environmental metrics are meant to represent a simplified version of reality from a series of observations (Eurostat 2014b). Thus, metrics provide information extending beyond that directly associated with the value of the parameter through which they are represented (OECD 1993). Because of the ability to approximate reality, environmental metrics act as boundary objects between science and policy thereby playing a key role in enabling evidence-based environmental governance. The term 'metric' is here used as an umbrella concept that encompasses individual indicators, indicator sets and composite indicators (i.e. indices). The differences between these concepts are described in Box 1.

Box 1: Metrics, indicators, indicators sets and indices

Metric: Indicator, indicator set or index.

Indicator: A parameter, or a value derived from parameters, which provides information about a phenomenon with a significance extending beyond that directly associated with a parameter value (OECD 1993).

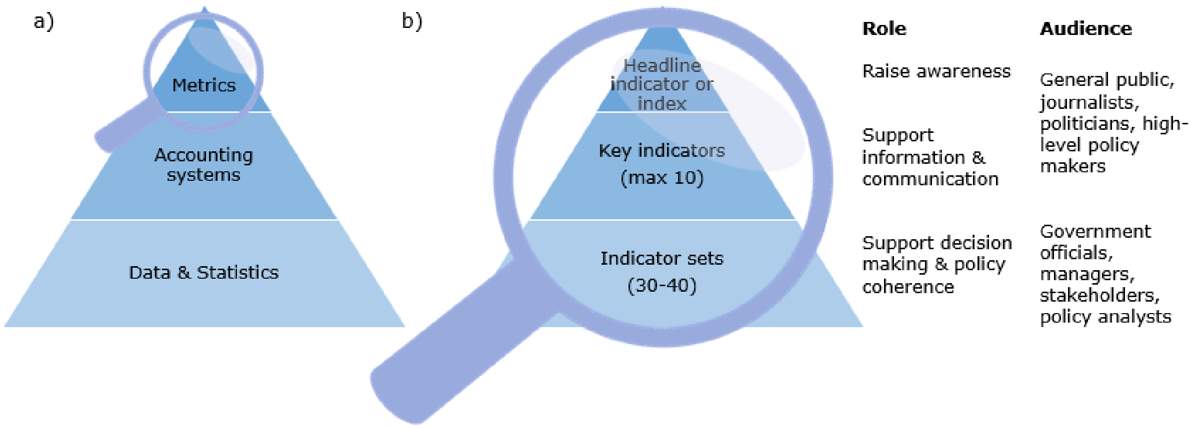
Environmental indicator: Indicators that have an environmental focus. They can commonly be grouped using the Driver-Pressure-State-Impact-Response (DPSIR) framework (EEA 1999, 2003b). In the DPSIR framework the environment is characterised through pressure (P), state (S) and impact (I) indicators. Pressures refer to anthropogenic factors such as emissions, physical and biological agents, the use of resources and land that act as stressors and therefore lead to changes in the state of the environment (EEA 2003b). State metrics provide a quantitative and/or qualitative description of physical (e.g. temperature), biological (e.g. fish stocks) and chemical (e.g. atmospheric CO_2 concentration) conditions in an area (EEA 2003b). In other words, they represent the biophysical conditions of the environment. In this thesis, they also describe social conditions related to environmental topics (e.g. human health related to outdoor air pollution, access to safe drinking water, etc.). Changes in state affect the environmental functions provided by natural capital, which can at the same time result in changes in ecosystem services that benefit humans. Changes in both environmental functions and ecosystem services resulting from anthropogenic activities are characterised through impact indicators (Maxim et al. 2009). Drivers (D) represent human activities that exert pressures on the environment, while responses (R) refer to actions taken to address the main drivers or to alleviate the burden on the environment or human health.

Indicator set: A list of indicators selected based on a common policy or conceptual framework.

Index: Composite measure that provides a single aggregated score of a phenomenon of study. It generally implies converting a set of indicators to common units (or a unitless scale) and assigning weights (i.e., averaging, adding, or application of other mathematical operators) before aggregating them into a single score.

Environmental metrics can also be split in different categories depending on the intended use (Figure 1b). At the highest level, a single headline indicator, which sometimes takes the form of an index, is commonly used for awareness-raising purposes. This type of metric reduces an area of concern to a single aggregate value and thus can be used to reduce a complex problem to a simple idea, e.g. when the Gross Domestic Product (GDP) increases, the economy improves. Such reductionist approach has its benefits but needs to be supported by additional indicators that provide additional context. In this vein, small sets of key indicators can also be used to support information and communication in areas of public interest. Both headline metrics and key indicators are meant for non-technical audiences such as the general public, politicians or journalists. A more detailed analysis aimed at supporting the decision-making process demands a larger set of indicators, which commonly addresses a more technical audience, including policy-makers, policy analysists, stakeholders, etc.

Figure 1: Information and metrics



The figure on the left shows the information pyramid, while the figure of the rights shows different types of metrics depending on the purpose and target audience.

Source: Adapted from Eurostat (2014b) and OECD (2008)

Environmental metrics are at the core of policy making. The European Environment Agency (EEA 1999) describes four main uses of environmental metrics:

- provision of information on the state of the environment to support the evaluation of the urgency of environmental problems,
- identification of key factors behind environmental problems to support policy development, as well as priority and target setting,
- comparison of countries' performance over time,
- monitoring progress towards policy objectives and the effects of policies.

Whether it is to monitor sustainable development, environmental strategies or multilateral environmental agreements, many international and national organisations use environmental indicators to inform and monitor their policies, or to validate the narrative that underlies their vision. Yet despite the overwhelming scientific evidence that feeds into policy, there is a clear gap between the urgency of the environmental crisis and the policy measures put in place to tackle it.

One of the reasons for this gap is the limited practical impact the notion of environmental limits has had in policy. Although the concept has gained traction in recent years, it has mainly had a rhetorical – rather than an instrumental – use in policy (Pickering and Persson 2020). As a result, environmental policy targets continue to insufficiently weight scientific evidence of environmental degradation. After all, environmental target setting is a complex process in which besides environmental concerns, technological feasibility, economic consequences, distributional aspects, vested interests and other relevant factors are weighted (Moldan et al. 2012). National pledges for greenhouse gas (GHG) emission reduction falling short of meeting the global goal set in the Paris Agreement (UNFCCC 2015) is one of many possible examples of how the urgency of tackling environmental degradation is insufficiently weighted in policy responses. Given that the scientific evidence is unequivocal when it comes to stressing the urgency to act, one must wonder whether the way this information is translated into metrics can be a contributing factor to the implementation gap. In this context, it is fair to ask whether we are really measuring what matters.

1.2. The environmental dimension in economic welfare, sustainable development and environmental (sustainability) metrics

Indicators and indices reflect our interpretation of the phenomenon we intend to characterise. Sustainable development, economic welfare and environmental sustainability are no exception. The following subsections present some of the most well-known indicators and indices in these areas and describe their suitability to reflect the urgent environmental situation described by the scientific community.

1.2.1. Measuring sustainable development and sustainable economic welfare

The Brundtland definition of sustainable development[1] is still at the centre of the political discourse related to development and economic welfare, partly because of the openness of the concept, which has allowed different stakeholders to adapt it to their own contexts and purposes (Greco et al. 2019). In this vein, the concept has been broadly interpreted as non-declining or increasing human welfare. The actual implications of this interpretation led to a debate in economics around the factors that contribute to human welfare and their substitutability. It is commonly accepted – at least among ecological economists – that there are different types of capital that contribute to human welfare: manufactured, social, human and natural capital (Ekins 1992). Whether the contributions of natural capital are unique or can be replaced by those provided by other forms of capital is at the core of the concepts of weak and strong sustainability. In short, weak sustainability assumes that welfare depends on an aggregate stock of capital that is independent from the type. Thus, under this proposition, the functions provided by natural and manufactured capitals are interchangeable. On the other end, strong sustainability considers that the substitution of natural capital by other types of capital is limited because certain elements of natural capital provide unique and irreplaceable functions. In this line, from a strong sustainability perspective, development should ensure that the unique functions provided by natural capital are sustained over time, irrespective of those of manufactured, social and human capital (Ekins et al. 2003a). More details on this are provided in chapter 2. Following these propositions, there are different ways in which natural capital is represented in the most prominent sustainable development and economic welfare metrics.

The first category covers weak sustainability metrics that take the form of macro-economic aggregates that include monetary measures relating to natural capital. The most prominent examples are the Genuine Progress Indicator (GPI) and Genuine Savings (GS) (Box 2). Both GPI and GS seem to be at odds with the insights provided by scientists on the state of the environment, partly because of methodological and data limitations related

[1] "Development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (Brundtland et al. 1987).

to the valuation of natural capital (Ekins 2011), but also because they fail to highlight the urgency of current environmental challenges, including the activation of non-linear responses in some Earth System processes and the irreversible loss of some ecosystem services. For instance, the global per-capita GPI only slightly decreased since 1978 (Kubiszewski et al. 2013). Total wealth and wealth of natural capital, on the other hand, increased in the 1995-2014 period (Lange et al. 2018). In the meantime, the sixth mass extinction is underway (Ceballos et al. 2015) and several planetary boundaries have been breached (Steffen et al. 2015b). While monetary indicators such as those described above speak to stakeholders with mixed interests by bringing together environmental and socio-economic aspects in a single metric, they systematically fail to capture the trends described by biophysical indicators of the status of the environment.

Box 2: Monetary weak sustainability metrics

Genuine Progress Indicator: GPI corrects private consumption expenditures with the cost and benefits associated with manufactured, social, human and natural capital by using different valuation techniques (Kenny et al. 2019). With regard to natural capital, it incorporates the cost of different types of pollution (e.g. air, water), depletion of non-renewable resources, ozone depletion, climate change and loss of some ecosystems (Lawn 2003; Kubiszewski et al. 2013; Talberth and Weisdorf 2017), although the elements included differ depending on the entity for which GPI was calculated. These costs are commonly quantified through damage functions, except in the case of non-renewable resources, which uses a non-market valuation technique (Kenny et al. 2019). GPI assumes that the monetary losses attributed to the degradation of natural capital can be compensated by an increase in other types of capital from a current welfare perspective (Kubiszewski et al. 2013). Nonetheless, this would affect the capacity of economic welfare to be sustained over time. For this reason, GPI was never intended to be a measure of absolute sustainability and therefore is intended to be complemented by biophysical indicators (Kubiszewski et al. 2013).

Genuine Savings: The World Bank produces two related indicators: national wealth and GS (also referred to as adjusted net savings). The former is measured as the present and future value of the stock of manufactured, human and natural capital. Natural capital represents the discounted sum of the value of rents generated over the lifetime of fossil fuels, mineral resources, agricultural land, forests and protected areas. Maintaining or increasing total wealth over time is considered a criterion to ensure sustainable, long-term growth (Lange et al. 2018). GS, on the other hand, captures some of the factors that lead to changes in wealth over time. GS is measured as gross national saving minus depreciation of produced capital, depletion of natural capital, the cost of air pollution damage, plus a credit for education expenditures. While negative values indicate that a country is consuming more than it is saving – thereby jeopardising long-term sustainability – (Lange et al. 2018), positive values do not necessarily represent a sustainable trajectory if environmental and other externalities are not reflected in prices (Neumayer 2003).

A second category covers broader sustainable development metrics. Currently, the most prominent metrics in this category take the form of indicator sets that are centred around the Sustainable Development Goals (SDGs). These sets are used to measure both performance at a given year (OECD 2019; Sachs et al. 2019; UN 2019a) or progress over time (Eurostat 2019b; OECD 2019). The use of multiple indicators provides a more comprehensive picture than aggregates such as GPI and GS, but at the expense of increased complexity. For this reason, in some cases sustainable development indicators have been aggregated at SDG level (Eurostat 2019b; OECD 2019) or to create a single index (Sachs et al. 2019). Whether metrics in this group reflect weak or strong sustainability depends on several conditions that are described in the next section.

1.2.2. Measuring environmental sustainability through the lens of strong sustainability

While monetary aggregates of economic welfare are not suited to monitor environmental sustainability, it is unclear whether sustainable development metrics overall, as well as other environmental metrics can be used for that purpose. Understanding what environmental sustainability means in this context is the first step.

Environmental sustainability has been defined in different ways as shown in Table 1. Of course, these definitions are very broad, but they share some commonalities. A common theme of the definitions is that some features of natural capital need to be sustained indefinitely. Depending on the definition, these features are the stock of natural capital, its functions, or the benefits obtained therefrom. Given that abiotic resources cannot be replenished when using them, the stock of abiotic natural capital cannot be maintained indefinitely at any given level of use, and since the ability to provide benefits depends on the functioning of natural capital, it seems sensible to conclude that the focus should be set on maintaining the functions of natural capital. In this context, although the definition does not clarify which specific environmental functions need to be preserved or which particular elements of natural capital need to be targeted, it implies that some kind of 'sustainable' reference value is needed to indicate whether those conditions are met.

Table 1: Definitions of environmental sustainability

Source	**Definition**
Goodland (1995)	Maintenance of natural capital
Holdren et al. (1995)	Maintenance or improvement of the integrity of the life support system of Earth
Ekins et al. (2003b)	Maintenance of important environmental functions and therefore, the maintenance of the capacity of the natural capital stock to provide those functions.
Sutton (2004)	The ability to maintain the qualities that are valued in the physical environment
Moldan et al. (2012)	Maintaining nature's services at a suitable level

Thus, for a metric – whether an indicator or an index that aggregates indicators – to be able to reflect environmental sustainability, two conditions are proposed here. First, the indicators should be related to the status of natural capital and its capacity to function. Second, the indicators need to compare the current state of natural capital, or the pressure natural capital is subject to, to a reference situation that can be considered environmentally sustainable. Because most decisions related to environmental management and resource use are not made at the global scale, but rather at lower levels (Häyhä et al. 2016), a third criterion is added, which requires the indicator(s) to be relevant for nations.

Whether existing metrics meet these criteria has not been established. In order to shed light on this issue, a selection of well-known sustainable development and environmental metrics have been interrogated against these three criteria. These metrics include the various metrics based on the SDGs, environmental metrics such as the Environmental Performance Index, the Ecological Footprint or the indicator set in the Planetary Boundaries framework. The results of the assessment, which considers the general suitability of the metrics based on the methodological description material, are summarised in Table 2 and explained below.

Table 2: Overview of notable indicators related to the environment or environmental sustainability

Metrics	Type	Focus	Measures	Scale	References
SDG indicators [a]	Set	Environment	Performance against internationally agreed targets, best performing countries, or sustainability reference values	National and global	Eurostat (2019b); IAEG-SDGs (2019); OECD (2019)
SDG Index [a]	Composite	Environment	Performance against internationally agreed targets, best performing countries, or sustainability reference values	National and global	Lafortune et al. (2018)
Environmental Performance Index	Composite	Environment	Performance against internationally agreed targets, best performing countries, or sustainability reference values	National	Yale University (2018)
Ecological Footprint	Composite	Environmental sustainability at global level; self-sufficiency at national scale	Performance against countries' or Earth's regenerative capacity	National and global	Borucke et al. (2013); Lin et al. (2016)
Planetary Boundaries	Set	Environmental sustainability	Performance against environmental limits	Global	Rockström et al. (2009b); Steffen et al. (2015b)

[a]: Only the environmental indicators are considered.

The SDGs are the most recent policy-driven attempt to characterise the broader sustainable development concept (UN 2015a) after the mixed results of the Millennium Development Goals (UN 2015b). The SDGs comprise 17 headline goals and 169 targets which are monitored through 232 indicators, many of which have an environmental focus (ECOSOC 2018). Beyond the official indicator set, different institutions have adopted their own sets to capture their specific contexts (Eurostat 2019b; OECD 2019). In the official SDG set, most of the indicators with an environment focus do not represent natural capital and its functions, but other themes such as sustainable consumption and production, adoption of environmental policies and related mechanisms, or social aspects related to the environment (Campbell et al. 2020). As a result, most environmental SDG indicators are not suitable to represent environmental sustainability. Given that the adoption of alternative SDG indicator sets relies heavily on the official set, this also holds true for the sets adopted by Eurostat and OECD, as well as for the SDG Index. Additionally, while all these indicator sets and indices use reference values to contextualise country performance, these do not necessarily represent environmental sustainability conditions. Instead, the reference values used are a mix of internationally agreed targets, best performing countries and sustainability reference values. Consequently, environmental sustainability cannot be quantified systematically through SDG-based metrics (c.f. chapter 6).

The indicator set underlying the Environmental Performance Index faces similar shortcomings. In its latest edition (Yale University 2018), the index comprises 24 indicators arranged in ten issue categories and two policy objectives (environmental health and ecosystem vitality), which are then aggregated into a composite score at country level. The Environmental Performance Index, as the name clearly indicates, is an index of environmental performance, not environmental sustainability. While most indicators are related to the functions of natural capital, their performance is measured against a combination of international targets, best performers and sustainability reference values. Thus, the use of sustainability reference values is not widespread, thereby limiting the capacity of the index to monitor environmental sustainability.

The Planetary Boundaries framework defines safe boundaries for nine biophysical processes that regulate the stability of the Earth system (Steffen et al. 2015b). Distance from the boundary that indicates a 'safe operating space' is measured globally. Leaving aside scientific discussions around the existence or position of some boundaries, the framework uses sustainability reference values to measure performance. Nonetheless, the use of the indicators at lower scales is not very straightforward. So far, also, attempts to downscale the Planetary Boundaries framework to the national scale (Nykvist et al. 2013; Cole et al. 2014; Hoff et al. 2014; Dao et al. 2015; Lucas and Wilting 2018), which is the level at which most environmental policy is implemented, have limited consistency (Häyhä et al. 2016). While the Planetary Boundaries framework fulfils the first two criteria, its limitations at the national level hinder the suitability of the framework to monitor environmental sustainability at this scale.

Last, the Ecological Footprint addresses nations and uses a sustainability reference value, yet at this scale it is mainly used as an indicator of self-sufficiency, i.e. to show whether an ecological deficit exists at country level (Blomqvist et al. 2013b, 2013a; Rees and Wackernagel 2013). More importantly, its unit of measurement, the 'global hectare', is a complex hypothetical construct of doubtful scientific validity (Blomqvist et al. 2013b, 2013a; Giampietro and Saltelli 2014a, 2014b; van den Bergh and Grazi 2014, 2015), thereby limiting its credibility as indicator of environmental sustainability.

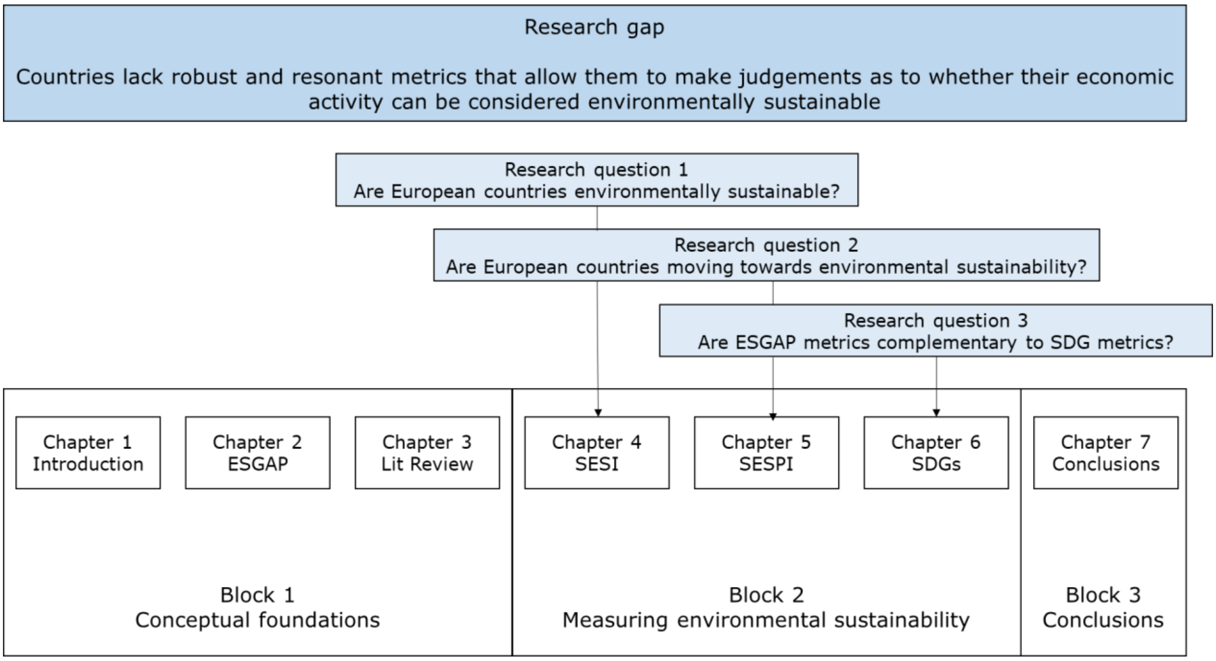
The existence of limits to the pressure humans can exert on the environment is widely acknowledged, but as this brief review shows, countries still lack robust and resonant metrics that allow them to make judgements as to whether their economic activity can be considered environmentally sustainable from a strong sustainability perspective. This is the research gap that will be addressed in this thesis. There are, of course, many more metrics that could have been included in the assessment, but the ones analysed above are among the most well-known and impactful ones, which increases their representativeness.

Ekins already identified this research gap two decades ago, which as shown above still remains, and developed the sustainability gap (SGAP) approach to address it (Ekins and Simon 1998, 1999; Ekins 2001; Ekins et al. 2003b). The approach described how absolute performance indicators across relevant environmental and resource issues could be combined to measure environmental sustainability and progress towards it. Back then, data availability only allowed to measure the performance of two countries against policy targets rather than environmental sustainability reference values (Ekins and Simon 2001). This thesis aims to update, improve, and operationalise the framework to address the research gap identified.

1.3. Research questions and scope of the thesis

The ultimate objective of this thesis is to improve the quantification of environmental sustainability at the national level as a way to provide more reliable information to decision makers. To that end, this thesis develops relevant metrics that allow measuring countries' performance against environmental sustainability criteria defined through natural and other relevant scientific disciplines, as well as monitoring progress towards or away from environmental sustainability. Such metrics represent a clear improvement over existing ones, which, as shown above, fail to capture environmental sustainability from a strong sustainability perspective The document is organised around three main blocks as shown in Figure 2.

Figure 2: Structure of the thesis



Full titles of the chapters: Chapter 1 Introduction: Introduction; Chapter 2 ESGAP: The Environmental Sustainability Gap (ESGAP) framework; Chapter 3 Lit Review: Science-based environmental standards; Chapter 4 SESI: Strong Environmental Sustainability Index; Chapter 5 SESPI: Strong Environmental Sustainability Progress Index (SESPI); Chapter 6 SDGs: Strong Sustainability and the Environmental Dimension of the SDGs; Chapter 7 Conclusions: Conclusions.

Block 1: Conceptual foundations

The first block provides the conceptual foundations on how to generate relevant metrics of environmental sustainability. Following this introduction, chapter 2 presents the Environmental Sustainability Gap (ESGAP) framework. ESGAP is the result of revising and further developing the original SGAP approach proposed by Ekins. The chapter reflects on which elements of the original approach have stood the test of time and which ones have prevented it from being implemented more widely. Based on that analysis, a renewed ESGAP framework is presented. ESGAP combines some of the elements of the original approach with new ones with the intention of facilitating its implementation. The renewed framework builds on already established concepts in ecological economics and environmental science such as strong sustainability, critical natural capital, environmental functions and science-based sustainability reference values. The renewed ESGAP also contains new indices of environmental sustainability that are computed at a later stage.

Sustainability reference values are at the core of environmental sustainability quantification. These reference values define the conditions for environmental sustainability as defined through natural science and other relevant scientific disciplines. Thus, chapter 3 contains a detailed account of the scattered literature that reports the progress made in producing the necessary scientific evidence to establish reference values against which the performance of countries can be measured.

Block 2: Measuring environmental sustainability

The second block represents the main contribution of this thesis to sustainability science. It implements the ESGAP framework to address the research gap highlighted above, i.e. the lack of appropriate metrics to monitor environmental sustainability. It consists of three chapters, each of which provides an answer to a research question.

Research question 1: Are European countries environmentally sustainable?

Chapter 4 computes the Strong Environmental Sustainability Index (SESI), one the new indices proposed as part of the renewed ESGAP framework. SESI is intended to measure the environmental sustainability of countries at a given point in time. The construction of the index follows the most comprehensive manual on composite indicators to date (OECD and JRC 2008). During its construction, methodological choices have been aligned to the extent possible with the theoretical framework described in chapter 2, thereby increasing the conceptual consistency of the final product. The index is flexible in that the underlying indicators can satisfy the information needs of researchers, statisticians and policy analysist, while the aggregated scores can be used to summarise the big picture to the general public or politicians that commonly require simpler, more condensed and easier to interpret information (Janoušková et al. 2018).

At this point, it is not possible to compute these indices for all the world's countries due to insufficient data. Nonetheless, the European Environment Agency and its European Topic Centres, the Joint Research Centre of the European Commission and Eurostat produce a wealth of environmental data and indicators that can be used to compute these metrics for European countries in a comparable manner (although with caveats). For this reason, the assessment is restricted to EU27 Member States plus United Kingdom (hereinafter Europe, European countries and European block for readability purposes).

Research question 2: Are European countries moving towards environmental sustainability?

Chapter 5 computes the Strong Environmental Sustainability Progress Index (SESPI), the second new index in the ESGAP framework. As opposed to SESI, SESPI measures progress towards or away from environmental sustainability, thereby providing a complementary perspective to the snapshot view of SESI. Its goal is to make the information on trends more digestible to different audiences. As with SESI, key methodological choices have been aligned with the ESGAP framework. SESPI is also computed for European countries due to data availability issues.

Research question 3: Are the ESGAP metrics complementary to SDG-based metrics?

Chapter 4 and chapter 5 compute new metrics of environmental sustainability that cover a specific niche in sustainability science. Although they are conceptually superior to other metrics when it comes to measuring environmental sustainability, there are many incumbents that measure related concepts such as sustainable development or environmental performance. In this context, it is important to understand how ESGAP metrics overlap with and can complement other metrics that are widely used. Specifically, chapter 6 elaborates on the value added of the ESGAP metrics based on a qualitative comparison with well-known metrics of sustainable development such as the SDG Indicators and the SDG Index. This ultimately shows whether the indices from the ESGAP framework provide information that is not captured by the SDG-related metrics and therefore, if it can be used to provide complementary narratives focused on strong sustainability.

Block 3: Conclusions

Chapter 7 is the final chapter in this thesis. It summarises the main findings and describes the research and policy implications of the research. Likewise, it provides an outlook of how this research could be expanded in the future.

2. The Environmental Sustainability Gap framework

2.1. Background

The SGAP approach was developed already in the late 1990s to respond to the metric gap described in the previous section (Ekins and Simon 1999). SGAP described how to measure absolute performance and progress towards environmental sustainability through indices that could be easily communicated to high-level policy makers and the general public. The approach builds on the concepts of critical natural capital – natural capital that performs important and irreplaceable functions – (Ekins et al. 2003a) and strong environmental sustainability, which assumes limited substitution capacity between natural capital and other types of capital, as well as between the diverse functions of natural capital (Ekins and Simon 1999). Building on those concepts, SGAP defined environmental sustainability and the criteria that can be used to characterise it. Although highly cited, the approach was only operationalized once because of lack of adequate data (Ekins and Simon 2001).

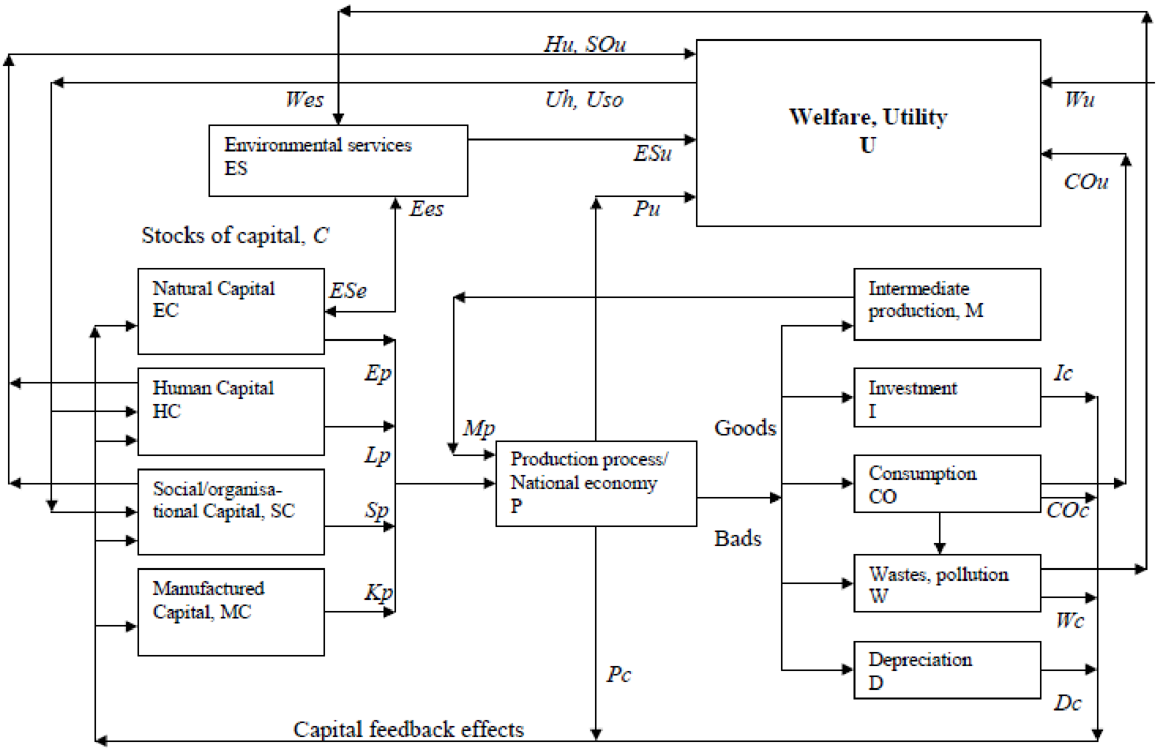
Various elements that were part of the thinking behind the SGAP approach have been widely embedded in contemporary policy making, as can be illustrated through a number of examples. Most obviously, the 1.5-2ºC targets in the Paris Agreement under the United Nations Framework Convention on Climate Change seek to maintain the essential functions of climate stability; the provisions in the Montreal Protocol to reduce the emissions of ozone-depleting substances (ODS) – eventually to zero – were driven by the scientific requirements to close the hole in the stratospheric ozone layer. The Oslo Protocol to the Convention on Long-Range Transboundary Air Pollution adopted the critical loads approach, such that emission reductions were determined according to "a quantitative estimate of an exposure to one or more pollutants below which significant harmful effects on specified sensitive elements of the environment does not occur according to present knowledge" (UBA 2004), which is clearly related to the maintenance of environmental function. Regulations in the European Union (EU) that limit exposure of humans to air pollution are informed by the World Health Organization's estimates of levels that will not harm human health. The EU's Water Framework Directive sets its objectives in terms of achieving and then maintaining 'good status' of water bodies, defined such that "[t]he values of the biological quality elements for the surface water body type show low levels of distortion resulting from human activity, but deviate only slightly from those normally associated with the surface water body type under undisturbed conditions (European Parliament and European Council 2000). Likewise, the levels of pollutants in accordance with 'good status' are required to be within "the range established so as to ensure the functioning of the type specific ecosystem and the achievement of the values specified above for the biological quality elements". Where human health is concerned, further regulations are set to ensure safety of, for example, drinking and bathing waters (European Council 1998; European Parliament and European Council 2006). In all these areas, policy making has built on science-based reference values following a desire to maintain environmental functions at a level that will ensure ecosystem and human health. These policy approaches reflect strong sustainability thinking, that does not seek to trade off environmental functions for perceived economic or social benefits, and aims to maintain critical natural capital because of a perception that it delivers goods and services that can be provided by other forms of capital only more expensively, or less adequately, or not at all.

The original SGAP approach did not derive a full set of indicators that would enable policy makers at the national level to have a comprehensive view of the extent to which environmental sustainability was being achieved across the full range of environmental issues. This has now been achieved with the ESGAP framework described in the next sections, the essential building blocks of which will now be briefly reviewed. Thus, sections 2.2, 2.3 and 2.4 summarise and develop further key concepts of the framework largely building on previous work by Ekins (Ekins and Simon 1999, 2001; Ekins et al. 2003b). Section 2.5 elaborates on the different reference values that can be used to characterise environmental sustainability. The revised ESGAP framework comprises three main metrics of environmental sustainability: SESI, SESPI and monetary environmental sustainability gap. All these metrics are further described in section 2.6.

2.2. Strong sustainability

Human well-being rests on the combination of different types of capitals. Ekins (1992) proposed a four-capital model in which natural, manufactured, human and social capital are combined to generate welfare as shown in Figure 3. Broadly speaking, the stocks of the four types of capital produce flows of services that feed into a production process that generates goods and services. These goods and services contribute to welfare in different ways. As explained in Ekins et al. (2019) (and previously developed in more detail in Ekins (2000)), the role of natural capital in welfare creation goes beyond its contribution as an input in the economic system, since it also contributes through the provision of services of a non-economic nature. On the other side, the system leads to 'bads' in the form of depreciation, and pollution and wastes, which affect negatively the capital stocks, and which need to be compensated for by investment if the level of the stock is to be maintained.

Figure 3: The four-capital model of welfare creation



Source: Ekins (2000)

The substitutability of the different types of capital has been largely debated, especially in the context of natural capital. This is at the core of the weak vs strong sustainability concepts (Costanza and Daly 1992; Neumayer 2003). The proponents of weak sustainability assume that welfare does not depend on a given type of capital, but on the aggregation of all of them (e.g. Genuine Progress Indicator (Kubiszewski et al. 2013) and Adjusted Net Savings (Lange et al. 2018)), thereby implying that one type can replace the other, although with exceptions. Strong sustainability, on the other hand, assumes that there is limited substitution capacity between different types of capital. In particular, the substitution of the functions provided by natural capital are limited by characteristics such as irreversibility, uncertainty and the existence of 'critical' components of natural capital, which make a unique contribution to welfare (Costanza and Daly 1992). Within natural capital itself, the functions provided by specific elements cannot be commonly replaced by those provided by other elements either (Neumayer 2003). The issue of substitutability has implications beyond the measurement of welfare, since it fixes a position on acceptable natural capital depletion and degradation (Barbier and Burgess 2017).

Although often presented as fixed positions, some authors have further split these two categories based on additional degrees of substitutability, thereby giving rise to the following categories: very weak sustainability, weak sustainability, strong sustainability and very strong sustainability (Turner 1993). This allows viewing the weak-strong sustainability proposition not as an absolute dichotomy, but as a continuum where full and no substitutability are the ends.

2.3. Natural capital

Natural capital represents "the elements of nature that directly and indirectly produce value or benefits to people, including ecosystems, species, freshwater, land, minerals, the air and oceans, as well as natural processes and functions" (NCC 2014, p.21). The benefits provided by natural capital range from the basic processes that regulate the Earth System, to goods such as food or fresh water that are indispensable for our subsistence, or the materials that represent the physical foundations of our infrastructure.

The stocks or assets of natural capital fulfil different types of functions that ultimately define their capacity to provide ecosystem services (flows). These functions are a subset of the physical, chemical or biological interactions between the components and processes of ecosystems (de Groot et al. 2010). Flows of ecosystem services, on the other hand, represent the "direct and indirect contributions of natural capital to human well-being" (de Groot et al. 2010, p.25). Often the environmental function (the capacity to provide a good or service) is essentially identical to the good or service itself, e.g. the service of providing air compatible with good health from breathing depends on the capacity of the environment adequately to disperse or otherwise remove pollution in a given location. In what follows, therefore, the environmental function may be indistinguishable from the good or service to which it gives rise.

The functions of natural capital may be seen as being of four broad kinds (Ekins and Simon 2003), although other classifications exist (e.g. de Groot et al. (2002)):

- *Source functions* represent the capacity of natural capital to sustain the supply of resources and therefore cover the provision of different type of resources used by humans, which include the formation of topsoil, the provision of space for human activities, the supply of water, minerals, fossil fuels biomass, etc.
- *Sink functions* represent the capacity of natural capital to neutralise wastes without incurring ecosystem change or damage. This includes the regulation of the chemical composition of the atmosphere and oceans and the assimilation of waste.
- *Life support functions* refer to the capacity of natural capital to maintain ecosystem health and function, which covers functions from the provision of quality habitat to the regulation of runoff and climate or the maintenance of biodiversity.
- *Human health and welfare functions* represent the capacity of natural capital to provide other services to humans, very often of a non-economic kind, which maintain health and contribute to human well-being in other ways. These could be related to amenity as in sites that have aesthetic, spiritual, religious or scientific value, or the capacity to provide space for recreation.

The functions are clearly inter-related. For example, the operation of both the source and sink functions are clearly important for the life support functions, and all three of these types of functions can affect human health and welfare. But, as shown in Table 3, the indicators that show the operation of these functions can be made distinct according to this typology.

Table 3: Classification of environmental functions by type of natural capital

Type	Life support (LS)	Sink (Si)	Source (So)	Human health and welfare (HW)
Air (includes atmosphere, outer space)	1.1LS Fulfilment of habitat air requirements (quantity and quality)	1.5Si Regulation of the chemical composition of the atmosphere	1.7So Oxygen 1.8So CO_2 (for plants)	1.9HW Air for respiration 1.10-1.14HW Aesthetic, spiritual, religious, historic

	1.2LS Protection against harmful cosmic influence 1.3LS Regulation of the local and global energy balances 1.4LS Regulation of the local and global climate (inc. the hydrological cycle)	1.6Si Dispersion and dilution of air emissions		(heritage value), scientific and educational information, cultural and artistic inspiration
Water (includes fresh and sea water)	2.1LS Fulfilment of habitat water requirements (quantity and quality) 2.2LS Regulation of runoff and flood protection (watershed protection)	2.3Si Regulation of the chemical composition of the oceans 2.4Si Dispersion and dilution of emissions to water	2.5So Water catchment and groundwater recharge 2.6So Water (for drinking, irrigation, industry etc.) 2.7So Medium for transport	2.8HW Purification of water for human consumption 2.9HW Provision and purification of water for recreation 2.10-2.14HW Aesthetic, spiritual, religious, historic (heritage value), scientific and educational information, cultural and artistic inspiration
Land (including soil, space, landscape)	3.1LS Providing fertility for habitats and ecosystems 3.2LS Providing space for habitats and ecosystems 3.3LS Climate regulation by means of carbon storage	3.4Si Containment of emissions to land 3.5Si Decomposition, dispersion, and dilution of emissions to land	3.6So Formation of topsoil and maintenance of soil fertility 3.7So Mineral resources for construction, industrial, commercial and ornamental use 3.8So Fossil fuels 3.9So Providing space for human habitation, transport, agriculture, other economic activities	3.10HW Providing space for recreation 3.11-3.15HW Aesthetic, spiritual, religious, historic (heritage value), scientific and educational information, cultural and artistic inspiration
Habitats (including ecosystems, flora and fauna, biomass)	4.1LS Storage and recycling of organic matter 4.2LS Storage and recycling of nutrients 4.3LS Regulation of biological control mechanisms 4.4LS Maintenance of migration and nursery habitats 4.5LS Maintenance of biological and genetic diversity	4.6Si Storage and recycling of human wastes	4.7So Prevention of soil erosion and sediment control 4.8So Fixation of solar energy and biomass production 4.9So Energy conversion 4.10So Biomass for terrestrial or marine foods and drinks, genetic and medicinal resources, biochemicals, fuel, fodder, fertiliser, construction, clothing and household fabrics, and ornaments	4.11HW Nature protection 4.12-4.15HW Aesthetic, spiritual, religious, scientific and educational information, cultural and artistic inspiration

Source: Slightly adapted from Ekins and Simon (2003)

Not all the desired uses of environmental functions are consistent or possible, for natural capital and its functions are scarce goods. Thus, they may be seen as if they compete with other in some cases. For example, a lake may fulfil different functions, e.g. it can be a source of fish, drinking water or irrigation water; sink of human or industrial waste; habitat for fauna and flora (life support); and serve as a place for swimming/sailing (human health and welfare), yet the use of one function may rule out or compromise the delivery of another from the same resource (Hueting 1980). Competition between functions can take different forms. It can be quantitative when one use precludes another and leads to depletion, e.g. extraction of non-renewable materials such as fossil fuels or industrial metals. Likewise, competition can be qualitative when one use reduces other functions. For instance, the provision of breathable air in urban environments (source) is impaired when urban air acts as sink of air pollutants, which ultimately damages human health, plants and buildings. Competition also has a spatial dimension if one use limits or precludes another through congestion, e.g. when deciding to clear a forest that contributes to carbon storage (sink) and the maintenance of biodiversity (life support) in order to expand agricultural land for food production (source). Hence, only by considering the full range of human impacts on the functions of natural capital can the latter be managed sustainably.

2.4. Environmental sustainability

Environmental sustainability has been defined as "the maintenance of important environmental functions and therefore, the maintenance of the capacity of the capital stock to provide those functions" (Ekins et al. 2003b, p. 612). The definition suggests that environmental sustainability should be represented through biophysical indicators, but leaves open two key issues: which specific functions need to be maintained and which level would ensure their maintenance in the long-term.

From this definition what matters about the environment is not particular stocks of natural capital per se, but the ability of the capital stock as a whole to be able to continue to perform the environmental functions which make – directly or indirectly – an important contribution to human welfare. In a situation of complete knowledge about the contribution of different functions to human well-being, their importance could be evaluated in these terms and the functions thereby deemed to be of high importance related back to the particular stocks of environmental capital which are responsible for them. De Groot et al. (2003) and Brand (2009) proposed several criteria to identify such 'critical natural capital' based on its importance and the threat level natural capital is subject to. Despite the considerable progress that has been made in understanding the contributions of natural capital to human well-being (Millennium Ecosystem Assessment 2005; Díaz et al. 2018), there is still enormous uncertainty about associated with the identification of all the functions that need to be maintained in different social contexts and geographical scales. In the absence of such information, it seems preferable to identify as 'important', or critical (and therefore essential for environmental sustainability), any environmental function that cannot be replaced by any other function, or the loss of which would be irreversible and (potentially) lead to immoderate costs and impacts on human health and welfare.

Against this background, Ekins and Simon (1999) proposed, building on the work of Daly (1991) and Turner (1993), a set of general principles that could guide the management of natural capital stocks in a way that does not threaten their capacity to provide environmental functions (see Table 4). These principles require to ensure that renewable resources such as fish or forests are exploited at a level that allows them to be renewed over time, to exploit non-renewable resources at a rate that allows their future use, to keep pollution at a level at which ecosystems cannot neutralise it without incurring in excessive damage, to maintain the capacity of ecosystems to support life, to respect human health standards and to conserve the elements of natural capital that provide additional services to humans. The precautionary principle governs the other principles, especially in the cases where uncertainty and the potential damage from the loss of functions, as in the case of life support functions, are higher.

Table 4: Functions of natural capital and environmental sustainability principles

Function	Objective	Principle	Description
Source	Maintain the capacity to supply resources	Renew renewable resources	The renewal of renewable resources must be fostered through the maintenance of soil fertility, hydrobiological cycles and necessary vegetative cover and the rigorous enforcement of sustainable harvesting. The latter implies basing harvesting rates on the most conservative estimates of stock levels for such resources as fish; ensuring that replanting becomes an essential part of forestry; and using technologies for cultivation and harvest that do not degrade the relevant ecosystem and deplete neither the soil nor genetic diversity.
		Use non-renewables prudently	Depletion of non-renewable resources should seek to balance the maintenance of a minimum life-expectancy of the resource with the development of substitutes for it.
Sink	Maintain the capacity to neutralise wastes, without incurring ecosystem change or damage	Prevent global warming, ozone depletion	Anthropogenic destabilisation of global environmental processes, such as climate patterns or the ozone layer, must be prevented.
		Respect critical levels and critical loads for ecosystems	Emissions into air, soil and water must not exceed their critical level and/or critical load, that is the capability of the receiving media to disperse, absorb, neutralise and recycle them, without disturbing other functions.
Life support	Maintain the capacity to sustain ecosystem health and function	Maintain biodiversity and ecosystem health	Critical ecosystems and ecological features must be absolutely protected to maintain biological diversity, which underpins the productivity and resilience of ecosystems.
Human health and welfare	Maintain the capacity to maintain human health and generate human welfare in other ways	Respect standards for human health	Emissions into air, soil and water must not exceed dangerous levels for human health.
		Conserve landscape and amenity	Natural capital elements of special human or ecological significance, because of their rarity, aesthetic quality, recreational values or cultural or spiritual associations, should be preserved.

Source: Adapted from Ekins and Simon (1999); Ekins et al. (2003b)

2.5. Environmental sustainability reference values

In order to make the above sustainability principles operational, quantitative sustainability reference values need to be defined against which current environmental states, pressures or impacts may be compared. Here we distinguish three types of sustainability reference values: environmental limits, environmental standards and environmental targets, although other typologies exist (Moldan et al. 2012; Vea et al. 2020).

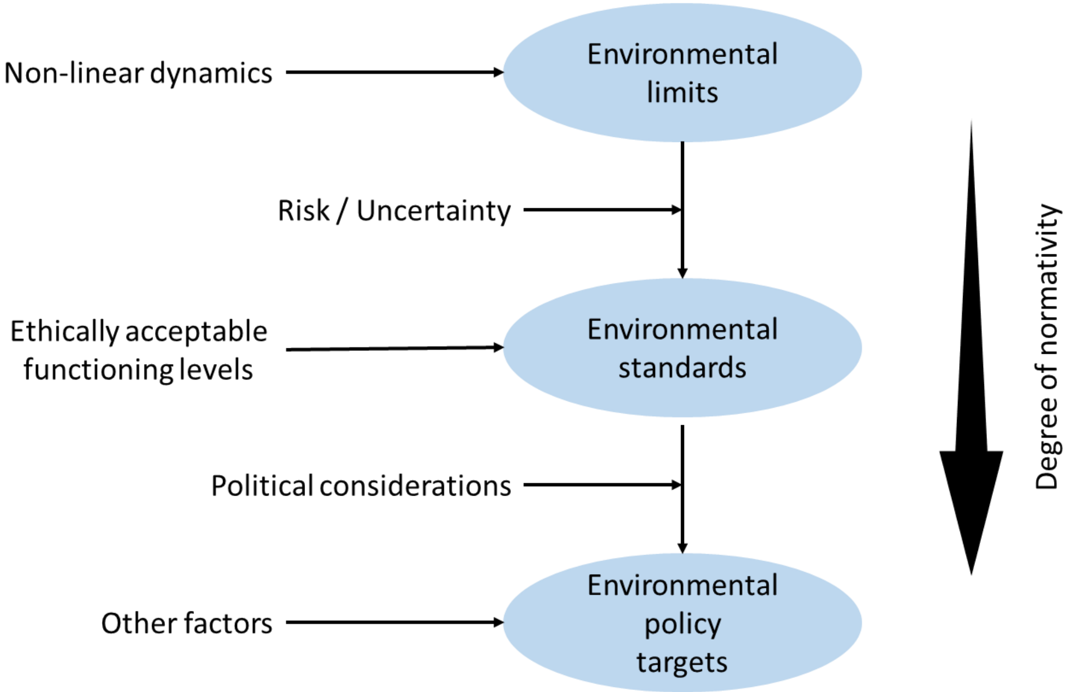
An *environmental limit* represents a point beyond which non-linear dynamics significantly change the functions and/or structure of an ecosystem. Non-linear dynamics describe the process by which a small pressure change leads to a disproportionate ecological response (Capon et al. 2015), which sometimes can result in a regime shift (Biggs et al. 2018). Examples of regime shifts include the collapse of fisheries (Bavington 2010), algae blooms in lake ecosystems (Carpenter et al. 2007), transitions from forest to savannah (Lovejoy and Nobre 2018), and many more. Some of these processes, especially those that are of global nature or have global implications are at the core of the Planetary Boundaries framework (Rockström et al. 2009a; Steffen et al. 2015b). In this context, it is worth noting that not all (eco)systems are subject to such behaviour (Schröder et al. 2005), as the sensitivity of ecosystems to pressures can vary greatly. There is some degree of normative judgement involved in the identification of environmental limits. Environmental sustainability reference values are considered limits when their transgression leads to non-linear dynamics that result in undesired consequences. What constitutes an undesired consequence may be significant deviations from natural conditions (e.g. from the natural variability of the Holocene climate (Steffen et al. 2015b)) or net losses in the provision of goods and services. Beyond those judgements, locating the position of the limit is a task for natural science. Nevertheless, limits are not universally fixed values, since the concrete position of a tipping point is influenced by other relevant biophysical parameters, the type of pressure, receptor or the resilience of the system itself (UBA 2004; Scheffer 2009; Bobbink and Hettelingh 2011).

Environmental standards are intended to depict the stock and quality of natural capital required to provide the necessary goods and services for society, while keeping a safe distance from environmental limits, taking account of the associated uncertainties. Like environmental limits, environmental standards are primarily science-based although value judgements are needed to define what a safe distance and acceptable service levels are. The decision in respect of the former depends on how society deals with risk and uncertainty, irreversibility and the threat of immoderate losses. There are different ways of defining an acceptable level of ecosystem goods and services. For instance, one could set such a level based on minimum material and emission requirements for a decent life (Steinberger and Roberts 2010; Di Giulio and Fuchs 2014), projections of future demand (Tilman et al. 2011; IRP 2019), health concerns (WHO 2000, 2005) or a range of ecosystem valuing techniques (de Groot et al. 2002). Once information on acceptable functioning levels is available, environmental standards can be determined based on the benefits-stock relationship that relates the ecosystem goods and services provided by natural capital to its quantitative and/or qualitative status. Likewise, the definition of some environmental standards also requires social norms such as ‘leave no one behind’ or ‘protect the vulnerable’ when dealing with issues such as access or impacts on humans.

Environmental policy targets usually deviate from science-based environmental standards, as the adoption of targets is the result of weighing not only environmental concerns, but also issues associated with technological feasibility, economic consequences and other politically relevant factors. As a result, environmental policy targets can become less stringent than environmental standards (Svancara et al. 2005; Doherty et al. 2018). Targets are derived mainly from policy documents and reflect people’s desires to the extent to which policies are aligned with social preferences.

Figure 4 summarises the relationship between environmental limits, standards and targets.

Figure 4: Relationship between environmental limits, standards and targets



Uncertainties in the identification of critical environmental functions are closely linked to the selection of environmental sustainability reference values. It can be argued that the life support and sink functions that are responsible for the regulation of the Earth System are among those that should be prioritised. These are the type of functions addressed in the Planetary Boundaries framework. However, the environmental sustainability concept used here is broader, for it also covers environmental sustainability at lower scales and incorporates economic and social aspects when these are associated with the exploitation of natural capital. Thus, environmental limits fall short from representing all the relevant functions of natural capital. Environmental standards, on the other hand, are more appropriate for such a task, although they also have limitations, e.g. higher degree of normative judgement. Environmental targets are also inadequate as a general rule. For environmental standards to become targets policy endorsement is needed, which is not always the case, as environmental targets usually represent a compromise between science, economic costs, social consequences and other relevant factors.

Environmental standards can take the form of 'sufficient' and/or 'necessary' conditions for environmental sustainability. The former describes the conditions that on their own are enough for the maintenance of a given environmental function. A necessary condition, on the other hand, represents a requirement that needs to be met, but that is not enough on its own (e.g. declining pressure without a specific 'sustainable' value). Ideally, environmental standards should represent sufficient conditions for environmental sustainability, but in cases when such standards are not available, necessary conditions might be used in order to include a relevant element of natural capital.

For life support and sink functions, renewable resources, and standards based on human health-related principles can be derived from natural and health sciences, although the knowledge base in each of these areas differs considerably. Functions related to maintaining a minimum life-expectancy of non-renewable materials or amenity are subject to broader social considerations. In all cases, standard setting leaves significant room for value judgements when defining the level at which environmental functions need to be maintained and/or how risk and uncertainty are dealt with. Such judgments are inevitably embedded in the environmental standards proposed by international institutions or scientists and therefore reflect their attitudes to risk.

There is no readily available dataset of environmental standards that can be used to operationalise the environmental sustainability principles presented in Table 4. Chapter 3 summarises the existing literature with the intention to select adequate environmental sustainability standards to calculate some of the headline metrics described in the next section.

2.6. Headline metrics of environmental sustainability

Three complementary metrics are proposed here as part of the ESGAP framework: SESI, SESPI and the monetary environmental sustainability gap. Beneath these three metrics, other composite indicators can be constructed, according to the typology of functions, or the principles of environmental sustainability set out above. This thesis focuses on SESI and SESPI, which are constructed following the guidance provided by the OECD manual on composite indicators (OECD and JRC 2008). Both indices are further described (and computed) in chapters 4 and 5 respectively.

2.6.1. Strong Environmental Sustainability Index

SESI provides a snapshot of a country's absolute performance against environmental standards that are linked to different environmental and resource areas. The indicators on which the index is based are intended to capture whether the capacity of natural capital to function is compromised over the long term. Each of the indicators is assigned a score between 0 and 100 based on a normalisation method, where 0 and 100 represent failure and compliance respectively with the environmental standard. In order to compute the final index, the normalised scores of the underlying indicators are aggregated across different layers, including the sustainability principles and the four functions presented above. A score of 100 indicates that the environmental standards of all the indicators are met. The difference between 100 and the index score would yield the physical environmental sustainability gap, the index previously proposed by Ekins and Simon (1999).

2.6.2. Strong Environmental Sustainability Progress Index

SESI provides a snapshot perspective on the functioning of natural capital. For this reason, Ekins and Simon (2001) proposed 'Years To Sustainability' as a second metric aimed at providing a general sense of whether a country was moving in the right direction from an environmental sustainability perspective. Based on linear trends, 'Years To Sustainability' showed the time it would take a country to meet all the environmental standards, which, although subject to strong assumptions, provided a clear and easy-to-understand message to policy makers. Nonetheless, this metric cannot be easily aggregated because negative trends in the individual indicators yield a score of infinity for 'Years To Sustainability'. As an alternative, SESPI is proposed here. SESPI shares the structure and underlying indicators of SESI. In order to capture the temporal dimension, two data points are used of each indicator to observed trends, similar to what Eurostat uses to measure progress towards the SDGs (Eurostat 2019b). Observed trends are compared to the ones that would be theoretically required to achieve the environmental standards at a given point in time, thereby giving a sense of whether enough progress is being made towards environmental sustainability.

2.6.3. Monetary environmental sustainability gap

The monetary environmental sustainability gap represents an aggregated monetary value of the maintenance costs (i.e. abatement, avoidance, restoration and protection costs) required to close the physical sustainability gap (i.e. the gap between sustainability conditions and SESI) for the relevant elements of natural capital, assuming previous losses are reversible. When divided by Gross Domestic Product, the resulting ratio is indicative of the 'unsustainability intensity' of the economy (Ekins 2001).

2.7. Differences between the SGAP approach and the ESGAP framework

All knowledge builds on previous knowledge. As such, the renewed ESGAP framework shares several elements with the original SGAP approach. At the core, both ESGAP and SGAP rely on the concepts of strong sustainability, critical natural capital, environmental functions, and science-based environmental standards. The first three concepts remain largely unaltered in the renewed ESGAP framework. The fourth concept, science-based environmental standards, has been made much more specific in previous sections with relevant conceptual clarifications in relation to its meaning and the differences with related terms such as environmental limits and environmental policy targets.

The indices to be calculated as part of the framework have also changed. In the original work of Ekins, environmental sustainability performance was measured through an index representing the 'physical sustainability gap'. Progress towards environmental sustainability, on the other hand, was calculated through 'years to sustainability'. In the ESGAP framework, both indices have been replaced by SESI and SESPI. A third composite indicator – the monetary environmental sustainability gap – remains unaltered.

The effects of these changes are most notable in the implementation of the framework. Ekins and Simon (2001) estimated the physical sustainability gap for seven environmental topics. For each of the topics, they calculated the difference between the situation in a given year and the environmental standard. One of the limitations was that it mainly used environmental policy targets as sustainability reference values, rather than science-based environmental standards. Thus, in practice, the authors measured a policy rather than a sustainability gap. A second limitation of the study was that the physical sustainability gap index lacked a coherent structure that could be linked to the theoretical framework. The study computed an index consisting of seven indicators that were not linked explicitly to the environmental functions and sustainability principles described in the approach taken. Likewise, the authors aggregated the indicator scores without any reflections on how the choices made during the construction of the index were related to the theoretical underpinnings of the index.

These issues have been addressed in this thesis, and in the following chapters in which the renewed ESGAP framework is implemented. Thus, ESGAP has been designed with the intention to facilitate the implementation of the original SGAP approach.

3. Science-based environmental standards

3.1. Background

Chapter 2 concluded that measuring the environmental sustainability of nations requires science-based environmental standards that are representative of the capacity of natural capital to function. This is a key distinguishing feature between the ESGAP metrics and other environmental and sustainable development metrics that use alternative reference values to contextualise the performance of countries.

The literature on environmental standards at the national level is very scattered. As a result, there is no readily available set of environmental standards that can be used to operationalize the ESGAP framework, although the recent review of approaches by Vea et al. (2020) is worth noting. The following sections present relevant reference values across a wide range of environmental and resource issues, many of which can be considered environmental standards. Most of them have been proposed in isolation, while others are part of wider frameworks such as Planetary Boundaries. Unequivocally determining whether a reference value is 'science-based' is not straightforward, as the term can be interpreted in different ways. Andersen et al. (2020) proposed three criteria to determine whether a reference value can be considered science-based: being achievable, being quantifiable, and being supported by a clear, analytical rationale. In practice, judging, for instance, whether a refence value is achievable requires a detailed knowledge of various disciplines that can only be obtained by bringing together several experts. Instead, the goal of this overview is to present an overview of reference values in the literature and to shortlist those that have a sufficiently clear and sound rationale to be considered in the indicator selection process in the next chapter.

Due to the number of topics addressed, this section only provides an overview of the main environmental standards proposed in the literature along with their rationale. This overview is structured around the environmental functions and sustainability principles described in Table 4. Given that an element of natural capital can be multifunctional, the allocation to the functions has been done based on the rationale of the environmental standards proposed. Nonetheless, full alignment was not always possible. The selection of the topics was also informed by the availability of relevant data in the indicators in chapter 4 in a back-and-forth process.

The overview presented here does not follow a rigid search strategy that is more characteristic of systematic reviews. Instead, it was undertaken in a more flexible manner that largely built on the snowball method once key references were identified. These key references differed depending on the topic addressed. In some cases, they were peer-reviewed papers, while in others they were grey literature and policy documents. This section covers environmental standards defined at country level, global standards that need to be downscaled and site-specific standards that need to be upscaled at country level. In doing so, the overview is focused on science-based reference values that can be used for European countries in line with the scope of this thesis.

3.2. Source functions

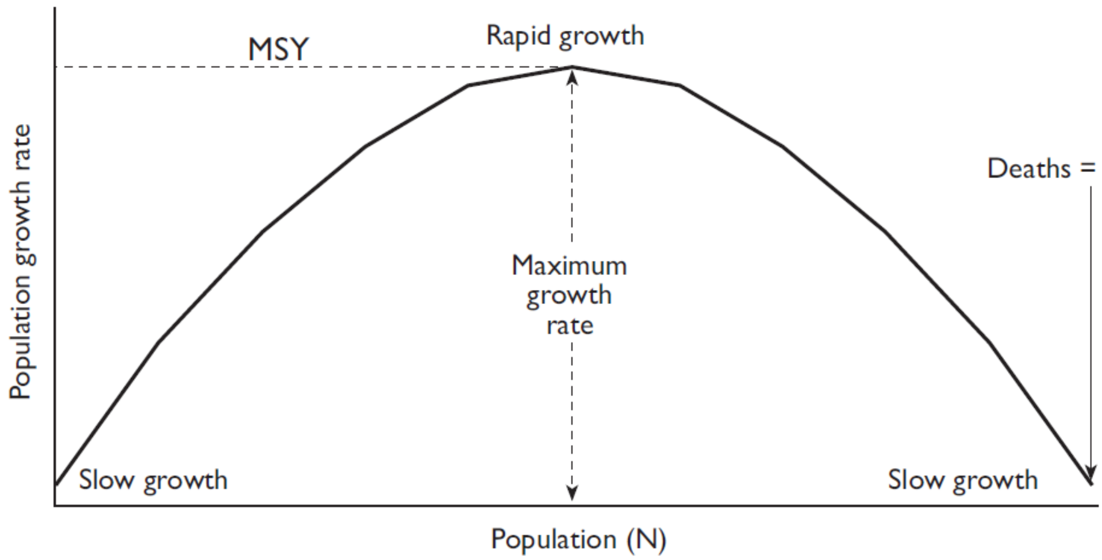
Source functions refer to the maintenance of the capacity to supply renewable and non-renewable resources and therefore usually take the form of exploitation rates. In the case of renewable resources, sustainable exploitation rates are based on the regenerative capacity of the resource. For non-renewable resources, on the other hand, scarcity is key and thereby sustainable exploitation rates are defined considering the exploitation potential of the resources over a given timeframe.

3.2.1. Renewable resources

3.2.1.1. Biomass from fish

Maximum Sustainable Yield (MSY) is a concept related to fish population management that is still used to estimate the maximum average biomass that can be harvested in the long-term without impeding the remaining stock in fisheries to reproduce itself (Bell and Morse 2008). In theory, MSY indicates the conditions that can sustain the maximum regenerative capacity of fish resources (Meltzer 2009) and therefore management practices around MSY require harvest rates to be adapted to the natural variability of stocks. MSY is commonly represented as the ratio between the fished and unfished stock. Different ranges can be found in the literature (Holt and Talbot 1978; Worm et al. 2009; FAO 2011), which are sensitive to the models used and the characteristics of the fish species, and therefore reflects the magnitude of the uncertainty of these estimates. Catch volumes that push the fish stock below that limit, and imbalances in the spawning population lead to the overexploitation and can jeopardise the future sustainability of the fishery. The most common graphical representation of MSY is shown in Figure 5. This considers a logistic growth function when biomass is unexploited.

Figure 5: The concept of maximum sustainable yield in harvesting a fish population



Source: Bell and Morse (2008)

Although conceived in the 1930s, the uptake of MSY did not occur until the 1950s when it became a target in fishery management (Punt and Smith 2002; Bell and Morse 2008; Meltzer 2009) following the development of various population dynamics models. Arguably the approaches to estimate MSY that have received greatest attention were those of Schaefer (1954) and Beverton and Holt (1957), which are also referred to as surplus production and yield-per-recruit models respectively (Punt and Smith 2002). These population models were intended to provide insights on how the stock of a given species would respond to certain management practices. The appeal of such methods was their simplicity and reasonable data requirements. Yet their main advantage was at the same time one of their main limitations, since the relatively simple mathematical representation of population dynamics contrasts with the complexity of the biology of fish species, especially when considering relevant factors such as competition, symbiotic or commensal relationships with other species, trophic relationships, or changes in carrying capacity due to pollution or other human influences (Holt and Talbot 1978). The adoption of MSY as a management target and the excessive level of institutionalisation of the concept led to prominent scientist to criticise its use (Larkin 1977; Holt and Talbot 1978).

Following the overexploitation of several fisheries in the 1970s, the management approach towards fisheries incorporated precautionary concerns and a more systemic view, and thus MSY is now considered a firm upper limit that provides a reference to measure overexploitation rather than a target in itself (Punt and Smith 2002; Worm et al. 2009). After all, despite its limitations, MSY is still considered to provide relevant information (Cochrane 2009; Sparholt and Cook 2010).

Because harvesting at MSY levels is not the goal, there are other meaningful concepts that are used to provide reference values in fisheries management. The most prominent ones are the spawning stock biomass that leads to low recruitment (B_{lim}) and the precautionary level that results in a reduced risk of low recruitment (B_{pa}) (ICES 2018). These reference values can take different forms. For instance, B_{lim} can represent the lowest observed spawning stock or the 'Minimum Biological Acceptable Level' (i.e. the spawning biomass level below which, observed spawning biomasses over a period of years, are considered unsatisfactory and the associated recruitments are smaller than the mean or median recruitment) (Cadima 2003). Accordingly, mortality rates F_{lim} and F_{pa} are meant to be consistent with B_{lim} and B_{pa}. Mortality rates below F_{pa} and spawning stock biomass above B_{pa} are considered signs of good status by the International Council for the Exploration of the Sea (ICES) (ICES 2003). The methods recommended by ICES to estimate the stock-specific reference values are documented in ICES (2003).

Currently, the main methods used in Europe to classify fish stocks assign an exploitation level based on status based on criteria of stock abundance, population age and size distribution, and reproductive capacity, although the specific indicators and reference values differ (EC 2010; FAO 2011). It is important to note that because fish populations compete between each other and because of predator-prey interactions, it is not possible for all fish populations to meet good quality standards at the same time (Piet et al. 2010).

3.2.1.2. Biomass from forestry

Forests are associated with a wide variety of ecosystem goods and services. Thus, they act as a source of materials and energy resources, support biodiversity through habitat provision, regulate water and temperature, maintain soil stability and fertility, store important amounts of carbon, etc. (Myastkivskyy 2012). Due to this multifunctionality, different criteria can be used to suggest environmental standards for forest conservation or exploitation levels. This section only considers environmental standards associated with the source function of forests.

Several studies have explored the maximum amount of wood resources that can be sustainably extracted from European forests (Forest Europe et al. 2007; EEA 2010; Forest Europe et al. 2011, 2015; O'Brien 2015). Utilisation rate is represented as the ratio between fellings and net annual increment (NAI), the latter being equal to gross increment minus natural losses (Tomter et al. 2016). The European Environment Agency esablished the sustainable utilisation rate at country level as 100% of NAI (EEA 2010), but this would lead to younger forests, lower biomass pools, depleted soil nutrient stocks and a loss of other ecosystem functions (Schulze et al. 2012). The current recommendation stands at 70% (EEA 2017). Using a unique relative exploitation rate for all countries could also mask relevant differences in the availability, potential and previous exploitation rates of forest resources that might justify setting different standards depending on regional characteristics. At the same time, while a 100% utilisation rate ensures that the net country stock is kept constant or increases, at higher geographical scales this should be seen as a firm upper limit considering the need for massive forest-based climate change mitigation (Smith et al. 2015).

3.2.1.3. Surface water

Water scarcity broadly refers to restricted water availability and can be defined as "a recurrent imbalance that arises from an overuse of water resources, caused by consumption being significantly higher than the natural renewable availability" (Strosser et al. 2012, p. 10). Beyond certain exploitation levels, water scarcity threatens the integrity of ecosystems that rely on freshwater, and poses limitations on economic and human activities.

From a human perspective, water scarcity occurs when there is not enough water of sufficient quality to meet human demands. These demands cover essential uses such as drinking, as well as uses related to agriculture, industry or energy cooling. Thus, characterising scarcity as a function of human demand has an inescapable social component both because demand is influenced by personal choices and policy, and because scarcity could also apply to situations where access to abundant water resources is lacking. The Water Stress Index is probably the most widely used scarcity indicator due to its relatively low data requirements and its simple message. It relates average human blue water demand to available blue water resources (Falkenmark 1989; Falkenmark et al. 1989) and provides the thresholds in Table 5 to characterise scarcity. In this case, scarcity compromises the uses beyond basic domestic water requirements, for the latter are quite small compared to the water demands from industrial and agricultural activities (Rijsberman 2006). The cases where households cannot satisfy their basic water demands are mostly related to lack of access to the resource, rather than to its absence (Rijsberman 2006).

Table 5: Reference values for Water Stress Index

Threshold	Condition
>1,700 m^3 per capita	Limited Stress
1,000-1,700 m^3 per capita	Stress
500-1,000 m^3 per capita	Scarcity
<500 m^3 per capita	Absolute Scarcity

Source: Based on Falkenmark (1989)

As pointed out by Damkjaer and Taylor (2017), the Water Stress Index was originally developed as an early-warning system related to food security for very specific geographical and climatic circumstances. One of the main limitations of Falkenmark's approach is that the reference values given to group communities' water scarcity levels do not reflect that water can be 'virtually' imported through traded goods and services (Hoekstra and Mekonnen 2011). This is particularly relevant in the globalization era.

A widely used alternative metric to characterise freshwater scarcity relates blue water demand to the resources available in a river basin. Such metrics have been referred to as Water Resources Vulnerability Index (Rijsberman 2006), use-availability ratio, withdrawal-to-availability ratio or criticality ratio (Hoekstra and Mekonnen 2011). Specifically, they represent scarcity as the percentage of total annual demand over available water resources. The European Environment Agency uses the so-called Water Exploitation Index (WEI) as a scarcity indicator. WEI is defined as the mean annual total blue water abstraction divided by the long-term average freshwater resources (EEA 2003a). These indicators are considered to give a rough estimate of pressures on water resources and the ecosystems they maintain (Szestay 1970; Raskin et al. 1997). Table 6 shows the use-availability reference values used to assess the condition of water bodies.

Table 6: Reference values for Water Exploitation Index

Threshold	Condition
<10%	No Stress
10-20%	Low Stress
20-40%	Stress
>40%	Severe Scarcity

Source: Based on EEA (2003a); Eurostat (2015)

EEA (2003a) and several authors such as Lutter and Giljum (2015) trace back the stress and severe scarcity reference values to Raskin et al. (1997) and Alcamo et al. (2000). At the same time, the Raskin et al. (1997) figures are based on those of Szestay (1970), and Falkenmark and Lindh (1976), both of which cite the work of Balcerski in 1963[2], who argued that a withdrawal-to-resource ratio beyond 20% would compromise European countries' economic development. Setting a single fixed reference value for all the countries only seems useful to raise a flag (Raskin et al. 1996), yet it is not clear to which extent the reference value given accurately represents stress problems. The 20% use-availability value has diffused through the work of different researchers and has rarely been disputed. Fifty years after it was presented in a workshop in Warsaw, the value is still used as a precautionary target. Alcamo et al. (2000) also took the Raskin et al. (1997) values as reference when examining potential future water scarcity around the world. In doing so, they also provided further detail for the upper end values of Table 6 by specifying the conditions 'high water stress' for a withdrawal-to-resource ratio between 40% and 80%, and 'very high water stress' for ratios above 80%.

[2] This reference could not be found on the Internet.

More recently, the European Environment Agency switched to Water Exploitation Index Plus (WEI+) to replace WEI (EEA 2016b). WEI+ uses consumptive water[3] as denominator and is calculated at river basin or lower scales on a three-month basis. WEI+ addresses some of the criticism around WEI, e.g. abstraction not being the best indicator to describe the pressure exerted on freshwater systems or that WEI usually disregards relevant spatial and temporal aspects, since it is usually presented at country level on an annual basis. Nonetheless, WEI+ still represents surface (including lakes) and groundwater (including fossil water reserves) together (Eurostat 2015). Hoekstra (Hoekstra et al. 2011; Hoekstra and Mekonnen 2011) also argued that comparing water consumption to actual runoff is problematic when runoff has decreased as a result of upstream water consumption within the basin or hydrological alterations. Thus, comparing water consumption to natural runoff would better represent scarcity.

Both WEI and WEI+ use the same reference values to represent scarcity (Table 6). These references can, to a certain extent, be seen as general rule-of-thumb values with, a priori, limited empirical validation, because 1) environmental flow requirements – i.e. the amount of water flows necessary to sustain aquatic habitats and relevant ecosystem process – are only considered implicitly, and 2) the same limits are adopted independent from whether water withdrawal or consumption is included in the numerator.

Another approach to set reference values is to explicitly consider the environmental flow requirements of river basins, yet over 200 different methods have been documented (Tharme 2003). The resulting environmental flow requirements can represent mean annual volumes or more commonly a combination of different monthly and event-based (e.g. low, intermediate and high flow seasons) allocations. Although existing methods are usually applied at river basin or lower levels, there is an increasing demand to undertake assessments at the global level. Given insufficient ecohydrological data, global assessments are mostly based on hydrological methods (Pastor et al. 2014). A few examples of environmental flow requirements are shown in Table 7.

Table 7: Environmental flow requirements from selected assessments

Environmental flow requirements	Method	Reference
75%, 60% and 45% of mean monthly flows for low-, intermediate- and high-flow months respectively	Hydrological	Steffen et al. (2015b)
25-46% of global mean annual flows with important variations between river basins and low- and high-flow seasons	A combination of hydrological methods	Pastor et al. (2014)
20-50% of mean annual flows	Hydrological	Smakhtin et al. (2004)

The use of hydrological methods to set rule of thumb environmental flow requirements has been criticised for not having a solid empirical basis (Arthington et al. 2006; Richter 2010). Arthington et al. (2006) also challenge the resulting figures for likely leading to severe environmental impacts. The use of holistic methods, which besides hydrological considerations, also take into account flow-response curves of the relevant biota, would increase the scientific credibility of the resulting environmental flow requirements and

[3] Water demand can take the form of abstraction or consumption. Abstraction indicates the volume of freshwater withdrawal from surface or groundwater bodies. Water consumption, on the other hand, makes reference to the volume of freshwater withdrawn that is then evaporated or incorporated into a product (Lutter and Giljum 2015). The difference between both metrics resides in non-consumptive water, i.e. water returned back to the basin from which it was withdrawn, which is part of water withdrawal, but not of water consumption metrics.

support more effective management practices. It should be noted that this type of model also relies substantially on expert judgement and tends to be expensive to apply. Until robust estimates are provided, Richter et al. (2012) advocates for the adoption of a presumptive standard (80% of mean daily flows as environmental flow requirements) based on the precautionary principle, yet many other low-cost methods exist as illustrated by Pastor et al. (2014).

Use-to-availability reference values and environmental flow requirements have been used to upscale scarcity-related sustainability standards. Rockström et al. (2009b) selected consumptive blue water as a proxy to control for blue and green water-related thresholds at regional and continental level that could potentially have effects at planetary scale. In doing so, they assumed 42,500 $km^3 yr^{-1}$ of blue water – 12,500-15,000 $km^3 yr^{-1}$ of which are currently accessible – and that physical scarcity levels are reached when consuming 40% of the available resources (5,000-6,000 $km^3 yr^{-1}$). Considering a 1,000 $km^3 yr^{-1}$ uncertainty in existing water withdrawal figures, they set a precautionary boundary focusing on accessible resources at the low end of the resulting range (i.e. 4,000 $km^3 yr^{-1}$ consumption of blue water). Gerten et al. (2013) revised Rockström's figure downwards (1,100-4,500 $km^3 yr^{-1}$) after explicitly considering environmental flow requirements (36-57% of mean annual flows based on the median and maximum values from five different bottom-up methods) and physical scarcity. As a result, in a revised version of the Planetary Boundaries framework, Steffen et al. (2015b) maintained Rockström's global figure (4,000 $km^3 yr^{-1}$) and added a regional boundary based on environmental flow requirements, which in this case were estimated to be 75%, 60% and 45% of mean monthly flows for low-, intermediate- and high-flow months respectively. More recently, Gleeson et al. (submitted) concluded that global blue water consumption is not an adequate metric to characterise the complexity and heterogeneity of the water cycle and its interactions with the Earth System at various time and space scales. The authors recommend developing control variables for different water stores (surface water, atmospheric water, soil moisture, groundwater and frozen water) that can properly characterise possible regime shifts that can affect the functioning of the Earth System.

Beyond blue water, Schyns et al. (2019) have quantified the maximum amount of green water that would be available for human use in order to set enough land aside for nature, which includes compliance with international targets on terrestrial protected areas.

3.2.1.4. Groundwater

The (often contested) concept of 'safe yield' has influenced the management of groundwater resources for many decades. Originally defined based on concerns around future supply, there is still the misconception that any abstraction below the natural recharge is sustainable (Molle et al. 2018). Nonetheless, the meaning of safe yield has evolved over time to incorporate water quality and other concerns (Alley and Leake 2004). Broader in scope, 'sustainable yield' refers to the level of exploitation that can be maintained over the long-term without unacceptable environmental, economic and social impacts (Alley and Leake 2004). Both terms are relatively vague, and thus, there are no agreed reference values that can help delineate the line between safe/sustainable and unsafe/unsustainable exploitation of groundwater bodies. Reference values used to characterise overexploitation would in any case be shaped by local conditions, for they depend on multiple factors such as recharge and discharge rates, water quality or the existence (and type) of groundwater-dependent ecosystems (Kalf and Woolley 2005).

In trying to translate normative concepts such as safe yield and sustainable yield into quantitative criteria, some authors have proposed environmental standards at the regional scale. For instance, informed by empirical evidence, Henriksen et al. (2008) argued that a maximum of 30% of groundwater could be abstracted in Denmark before water quality problems become apparent. This value was later revised to 20% or lower after considering the environmental water requirements of groundwater-dependent ecosystems (Henriksen et al. (2014) cited in Gejl et al. (2018)). In Australia, the values can range from 5 to 70% depending on the region and the approach adopted to define environmental standards (Murray et al. 2003). It should be noted that sustainable yield can be a problematic concept – even if defined in broad terms – in the case of non-renewable groundwater systems.

In Europe, the characterisation of the exploitation status of groundwater resources is regulated by the Water Framework Directive (European Parliament and European Council 2000). Groundwater bodies need to meet four criteria to be considered to be in good status. These cover requirements related to the abstraction-to-recharge ratio, impacts in surface water and terrestrial groundwater-dependent ecosystems, and saline and other intrusions resulting from changes in the flow direction (EC 2009).

3.2.2. Non-renewable resources

3.2.2.1. Soil

Due to the large amount of time required for soil formation, soil can be considered a non-renewable resource considering an average human life (FAO 2015). European policy documents have identified eight main threats to soils that can ultimately result in its loss or degradation (EC 2006). This section focuses on soil as a resource and therefore only addresses soil loss, including the loss of organic matter. Erosion, loss of organic matter, sealing and landslides have been highlighted as key threats leading to soil loss (EC 2006). Other threats such as soil pollution or biodiversity loss are addressed under different functions.

Soil erosion

Although erosion is a natural phenomenon, the soil loss rate is exacerbated by human action, especially agricultural practices. In a way, it could be said that we are mining the soil (FAO 2015). Erosion has been described as the most important threat to soil in Europe (Jones et al. 2004). Among the main acting forces behind it, water seems to be the most relevant one in Europe, although the contribution of wind is by no means negligible (Eckelmann et al. 2006).

Environmental standards for soil erosion are based on the concept of 'tolerable soil loss rate', which was developed in the 1940s and mainly used in the US. Attempts to establish tolerance rates mainly focused on the potential loss of productivity of soils. Originally, tolerance was understood as the soil loss rate that would allow the maintenance of the productivity of agricultural land over the long-term both from a physical as well as an economic perspective (McCormack et al. 1982). Under this perspective acceptable loss rates were defined assuming that above a minimum soil depth, fertilizers could compensate for the loss of soil (Johnson 1987). The values set by the United States Department of Agriculture (USDA) commonly ranged between 4.5 and 11.2 t ha^{-1} yr^{-1}

(Mannering 1981). Over the years, several researchers called for moving beyond agricultural productivity concerns to also include environmental criteria when defining the tolerable loss rates (Moldenhauer and Onstad 1975; Mannering 1981; Logan 1982) thereby addressing the other functions fulfilled by soil (e.g. pollution control, flood control, carbon storage). This seems still to be the predominant view today (Bazzoffi 2008; Li et al. 2009), although it remains underdeveloped (Verheijen et al. 2009). The most cited target is that of Morgan (2005), who estimated that erosion rates above 1 t ha^{-1} yr^{-1} could lead to water pollution by phosphorus.

Currently, the concept of the tolerable soil erosion rate is broader and linked to all the functions of soil. As defined by (Verheijen et al. 2009, p. 27), it refers to "any actual soil erosion rate at which a deterioration or loss of one or more soil functions does not occur". In practice, this definition is operationalised by assuming that when erosion rates are comparable to formation rates, soil functions are not compromised, although this hypothesis remains untested (Verheijen et al. 2009). Likewise, this approach implicitly assumes that natural soil erosion rates are equal to soil formation rates, which is not necessarily true, yet it still provides the most suitable basis to set tolerance rates that – in the absence of robust reference values from the previous approach – follow the precautionary principle (Verheijen et al. 2009). Since soil formation rates vary widely spatially depending on local characteristics such as climate, geology, soil type, topography, and vegetation; tolerable soil erosion rates do so as well.

In the US, the Environmental Protection Agency (US EPA 2003) slightly lowered the previous USDA rates to 2-11 t ha^{-1} yr^{-1}, with minimum rates for shallow soils with unfavourable subsoils and maximum rates for deep, well-drained productive soils. OECD (2013) also provides a range between 1-6 t ha^{-1} yr^{-1} where minimum and maximum values would refer to shallow sandy soils and deeper, well-developed soils respectively. In Europe, more restrictive tolerable erosion rates have been proposed. In this line, Jones et al. (2004) argued that erosion rates beyond 1 t ha^{-1} yr^{-1} are potentially irreversible in a time span of 50-100 years, which is in line with soil formation ranges in Europe (0.3-1.4 t ha^{-1} yr^{-1}) (Verheijen et al. 2009). Globally, Montgomery (2007) reported mean soil production rates of around 0.4 t ha^{-1} yr^{-1}. 1 t ha^{-1} yr^{-1} was also proposed by Huber et al. (2008) as a precautionary tolerable erosion rate, yet acknowledging that formation rates in some parts of Europe would justify a less restrictive limit. According to Verheijen et al. (2009), the 1 t ha^{-1} yr^{-1} standard is also consistent with limiting the impact of soil erosion/sediment production rates on water quality. Recently, Panagos et al. (2020) proposed 2 t ha^{-1} yr^{-1}.

Soil organic matter

Soil organic matter, which can also be expressed as soil organic carbon, is related to key biological (provision of substrate and nutrients for microbes), chemical (buffering and pH changes) and physical (stabilisation of soil structure) soil properties that ultimately determine the ability of soil to fulfil different functions (Krull et al. 2004). Soil organic matter is commonly considered a key indicator of soil quality (Gregorich et al. 1994; Reeves 1997; Bradley et al. 2004), although the general belief that 'the more, the better' also has its exceptions (Krull et al. 2004).

Environmental standards have been proposed based on the relationship between soil organic matter content and soil productivity (Janzen et al. 1992; Körschens et al. 1998) and structural stability (Kemper and Koch 1966; Greenland et al. 1975; Kay and Angers 1999). Nevertheless, posterior reviews concluded that the evidence that supports such standards is scarce (Loveland and Webb 2003; Huber et al. 2008; Schjønning et al. 2009). Environmental standards – if any – would depend on site-specific factors that affect soil properties and would differ between the different functions fulfilled by soil (Huber et al. 2008; Schjønning et al. 2009; Patrick et al. 2013).

Soil sealing and landslides

Soil sealing involves the conversion of rural land into the built environment. As a result, soil is irreversibly covered by completely or partly impermeable artificial material that interrupts the contact between the soil system and other compartments, thereby impeding processes such as infiltration, filtering of rainwater, evapotranspiration, geochemical cycles and energy transfers, etc. (Huber et al. 2008). Although the need to limit soil sealing is accepted, there is insufficient scientific evidence to set environmental standards. At this point, scientific expertise can only inform policy targets that aim at balancing the loss of environmental functions and socioeconomic development (Huber et al. 2008).

Setting environmental standard for landslides is also a challenging task. Any landslide will limit or threaten the functioning of the soil, but they often occur naturally sometimes exacerbated by climate change.

3.2.2.2. Abiotic raw materials

Scarcity is the key criterion to represent the source function in abiotic materials. Life Cycle Assessment is likely the field in which most efforts have been made to understand how the different dimensions of scarcity are translated into indicators. This has implications on the format environmental standards take.

Different indicators have been proposed in the context of Life Cycle Assessment, only some of which are related to the scarcity dimension of resources (Schulze et al. 2020). Other indicators consider other environmental functions of the sites from which abiotic resources would be extracted. In the context of the source function, only the capacity to provide resources is considered. Other functions are considered in their respective function category.

In Life Cycle Assessment, depletion of abiotic resources is commonly characterised using different forms of material-specific use-to-availability ratios (Guinée and Heijungs 1995; Drielsma et al. 2016), although other methods exist (JRC-IES 2011; Klinglmair et al. 2014; Alvarenga et al. 2016). The discussion around which availability indicator – reserves (economically mineable part of a measured resource), resources (material base with reasonable prospect for extraction), extractable global resources (crustal content extractable by humans) or crustal content (existing material base in the Earth’s crust) – better characterises the loss of the source function is still open. Data on reserves is subject to regular fluctuations due to the volatility of raw material prices. Although more stable, resource data also changes as investments in exploration increase and extraction technologies improve (Drielsma et al. 2016). Last, data on crustal content has remained largely constant across time, yet it provides limited policy-relevant insights on the loss of the source function in relevant timescales for humankind.

There have been limited efforts at identifying the extraction rates at which the source function would be maintained over an acceptable timeframe. Henckens et al. (2014) represent the most notable exception and uses extractable crustal content as reference. The authors argue that an extraction rate that could be maintained over 1,000 years could be considered sustainable, although they acknowledge the arbitrariness of the value. From this proposal and the Life Cycle Assessment indicators, it looks like scarcity of abiotic raw materials is considered a global issue rather than a national one. After all, the use-to-availability ratio reflects depletion or scarcity at the global level, since the availability value used is global. A similar ratio could be calculated for countries considering the domestic resource base and domestic extraction, but it is arguable whether considering the diverging resource endowments the preservation of domestic resources should be prioritised over global resources.

Economy-Wide Material Flow Analysis also covers abiotic (and biotic) materials, although it does it from an aggregated perspective, rather than from a material-specific perspective. Arguably, Material Flow Analysis indicators tend to reflect consumption, although reference values for consumption have been proposed based on extraction levels deemed sustainable. Nonetheless, 'sustainable extraction levels' have not been defined based on scarcity considerations, but rather, assuming that global extraction and consumption of materials can be considered a proxy for global environmental impacts. The reference values proposed in the literature are summarised in Box 3.

Box 3: Reference values for Economy-Wide Material Flow Analysis indicators

Economy-Wide Material Flow Analysis Indicators provide an overview of the inputs and outputs of the economy in aggregated terms. Many authors have proposed reference values for global extraction levels that are then translated into reference values for per capita consumption levels, which are more meaningful at the national level.

One of the first researchers to suggest material resource use targets was Schmidt-Bleek (1993), who took the stabilisation of the climate as the basis to propose targets for reducing global material extraction. The author put forward the target of reducing global extraction of materials by 50% in one generation. The 'factor 10' concept was also born in this context. In this case, industrialised countries would need to cut their GHG emissions by 90% if per capita emission were to converge in the future. Accordingly, Schmidt-Bleek also called for a comparable reduction of material consumption in industrialised countries. When proposing the 'Factor 4' concept, von Weizsäcker et al. (1998) did not challenge Schmidt-Bleek's global target. Instead, he focused on concrete examples that would contribute to doubling wealth while halving global resource extraction. The extent of the global dematerialisation needed was highly criticised by Kågeson (2000). He argued that the links between materials as a whole and GHG emissions is limited, which calls into question the validity of the target.

Building on the work of Schmidt-Bleek and von Weizsäcker, Bringezu (2009) also opted for maintaining the rationale of halving global resource consumption, but chose 2000 as the reference year. Nevertheless, considering the fast increase of global resource extraction since then, he then proposed to maintain global abiotic resource use at the level of the year 2000 instead of halving it (Bringezu 2011). In doing so, he did not argue against Schmidt-Bleek's rationale for not being valid from a scientific standpoint, but rather for not being realistic anymore due to the developments in global material use since the 1990s (Bringezu 2014).

Jäger (2014) also proposed to cap future global material use to match the extraction levels of beginning of the 21st century. Similar material consumption rates were proposed by others

(Schmidt-Bleek 2008; Ekins et al. 2009; IRP 2014), but as stressed by Bringezu (2015), they do not provide information on the indicator used. Other researchers such Dittrich et al. (2012) proposed to keep material extraction at the level of 1992, although they provided no basis to support that decision. Stricks et al. (2015), on the other hand, chose the 1970s as sustainable extraction levels given that it was only then that the ecological footprint of humanity surpassed the Earth's carrying capacity.

Bringezu (2015) later revised the approach for target setting, but as he highlighted, there "is still no hard scientific evidence of causal relationship between human-induced resource flows and the possible breakdown of life-supporting functions at continental or global scale from which those targets could directly be derived" (Bringezu 2015, p. 41). For this reason, the targets he proposed are still driven by the principles used in the previous paragraphs. Specifically, he proposed a corridor target for abiotic minerals in which the lower and upper ranges are consistent with halving and maintaining global extraction at the 2000 level respectively. Others have suggested targets for individual abiotic material categories such as fossil fuels (Mudgal et al. 2012; Jäger 2014), non-metallic minerals (Mudgal et al. 2012; Jäger 2014), or metals (Ekvall et al. 2015) for Europe, but unless relatable to a global 'sustainable' extraction value, these are of limited use. Further, except in the case of fossil fuels, the targets seem to be rather arbitrary.

3.3. Sink functions

3.3.1. Global processes

3.3.1.1. Climate change

Climate change has the potential to destabilise the functioning of the Earth system (Steffen et al. 2015b). The increase of average temperature is already leading to significant impacts on humans and ecosystems, some of which include extreme weather events, sea level rise, reductions of food yields, loss of ecosystems, etc. The intensity of the impacts is expected to increase with temperature, yet not all processes respond linearly (Schellnhuber et al. 2016).

Global targets have been proposed since the late 1980s based on the notion of unacceptable impacts. Accordingly, the need to prevent "dangerous anthropogenic interference with the climate system" became the main goal of the United Nations Framework Convention on Climate Change (UN 1992). Most global targets take the form of limits to global mean temperature increase, although there are some exceptions [4].

Limiting the increase of global mean temperature to less than 2°C compared to pre-industrial levels was the first major target proposed. This target ultimately became "a useful 'boundary object' interfacing between science, social science, and policymakers" for many years (Randalls 2010, p. 602). Although informed by science, the target is mainly influenced by value judgements and political considerations (Rockström et al. 2009b; Knutti et al. 2016), which poses a question on whether it can be considered a science-based environmental standard. The 2ºC target has been criticised for being irresponsible

[4] For instance, Rockström et al. (2009b) proposed 350 ppm of CO_2 and 1 W m^{-2} of radiative force change taking preindustrial levels as reference sustainability baseline. The authors stress that both reference values are considered compatible with the 2°C goal. The values have remained unaltered after the update carried out by Steffen et al. (2015b). Hansen et al. (2008) also proposed 350 ppm of CO_2 as an environmental standard based on paleoclimate data and the results of modelling exercises that included 'slow' climate feedback processes such as ice sheet disintegration, vegetation migration, and GHG release from soils, tundra or ocean sediments.

(Hansen 2005) and for mainly reflecting the concerns of high-income countries (Tschakert 2015). More than a decade ago Hansen (2005) argued that a 2°C increase would lead to unacceptable consequences and proposed to adopt a 1.7°C increase goal, which was then revised downward.

Previous to the Paris Agreement, more than 100 mid- and low-income countries pushed for a 1.5°C target due to the unbearable impacts a 2°C rise would pose on the most vulnerable countries such as small island states (Tschakert 2015) [5]. This apparent disagreement over the overall goal is reflected in the final text of the Paris Agreement, where the signing countries stressed the need to hold "the increase in the global average temperature to well below 2°C above preindustrial levels and pursuing efforts to limit the temperature increase to 1.5°C above preindustrial levels" (UNFCCC 2015, p. 22). Prominent scientists have argued that the goal adopted in the Paris agreement strike a balance between necessity and feasibility (Schellnhuber et al. 2016). Nevertheless, there are several tipping points that may have already been crossed and others that overlap with the temperature range of the Paris goal. These include the destruction of coral reefs, melting of Alpine glaciers and Greenland, the instability of the West Antarctic ice sheet, and ice-free Arctic summers (Schellnhuber et al. 2016). Recently, Sullivan et al. (2020) have shown that the capacity of tropical forests (especially in South America) to act as sinks could be impaired beyond a 2ºC global mean temperature increase. Although well established, the feasibility of global climate targets has been called into question given the speed at which the global economy needs to be decarbonised and the massive amount of negative emissions required until 2100 (Rogelj et al. 2015).

The focus on global targets has been criticised for allowing insufficient political action (Victor and Kennel 2014). Yet, translating the previous targets into carbon budgets that can be allocated to countries is not a straightforward task. Doing so requires converting global temperature targets into global carbon budgets and allocating the latter to countries.

The first step uses estimates of the transient climate response to cumulative emissions of carbon (Rogelj et al. 2016), which relies on the near linear relationship between cumulative GHG emissions and temperature change (Friedlingstein et al. 2014) to calculate the global average surface temperature change per unit of total cumulative anthropogenic CO_2 emissions. The estimation of global carbon budgets needs to consider factors such as CO_2 emissions, non-CO_2 GHG emissions, the cooling effect of aerosols, the warming effect of soot particles, feedback effects, etc. (WBGU 2009) along with the associated uncertainty (Peters 2016). Table 8 shows estimates of the remaining carbon budgets to meet the 1.5 and 2ºC targets with 33%, 50% and 67% probabilities.

Table 8: Remaining global CO_2 budget [Gt CO_2] for the 2018-2100 period to meet 1.5ºC and 2ºC targets without overshoot

	33% probability	**50% probability**	**67% probability**
1.5ºC increase	840	580	420
2ºC increase	2,030	1,500	1,170

Source: IPCC (2018)

[5] The differences over the impacts of the 1.5 and 2ºC targets have been described in detail elsewhere (IPCC 2018).

The second step requires allocating global carbon budgets to countries. There are several effort sharing principles (Höhne et al. 2014), all of which lead to very different results (van den Berg et al. 2019). In some cases, they even lead to negative emission budgets for industrialised countries with over-proportionate cumulative emissions to date. Although the historical responsibility of countries was acknowledged in the United Nations Framework Convention on Climate Change (UN 1992), there is no agreement on how the remaining carbon budget should be split between countries. What seems clear though, is that given current emission levels, only near-zero or negative emissions are sustainable over the long term.

3.3.1.2. Ozone layer

Stratospheric ozone destruction occurs as a result of a combination of specific climatic conditions and the presence of ODS. Increased emission of ODS during the second half of the 20th century, led to significant decreases of stratospheric ozone concentrations around the globe likely causing relevant impacts on both human health and ecosystems globally (UNEP 2010). This phenomenon is particularly acute in Antarctica where concentrations during the austral spring drop below 220 Dobson Units (DU) – the reference value considered as a 'hole' – in large parts of the continent.

The relationship between the concentration of ODS and stratospheric ozone destruction is of non-linear nature (Molina 2009). The global ODS concentration was around 2 ppb (in effective equivalent stratospheric chlorine, EESC) when the ozone hole first appeared over Antarctica in the late 1970s (Newman et al. 2007). Modelling results suggest that EESC concentrations of 30 ppb would be required for ozone reductions of the same magnitude to take place in the tropics (Newman et al. 2009). Nonetheless, impacts well below that point are considered unacceptable (Molina 2009).

The Montreal Protocol adopted in 1987 limited EESC concentration to 4 ppb, which led to a maximum extra-polar ozone loss of 5-6% (Molina 2009). The value is similar to the proposal by Rockström et al. (2009b). In this line, the authors proposed a maximum reduction in ozone concentration compared to the 1964-1980 levels[6], which is estimated to be 275 DU.

Global ozone standards have received little attention because of the success of the Montreal Protocol in bringing the emissions of ODS down. Although this is regarded as a key factor behind the early signs of recovery in Antarctica (Solomon et al. 2016), a rise in the emission of banned substances was detected and attributed to China (Montzka et al. 2018; Rigby et al. 2019), which may delay the recovery (Dhomse et al. 2019). This issue of rogue emissions seems to have been solved recently (Montzka et al. 2021).

In principle, long-term country commitments in the Montreal Protocol and its subsequent amendments can be broadly considered environmental standards, although more action is required to decrease the pressure on the ozone layer (EEA 2019a). European countries joined the Montreal Protocol as a block and therefore the targets refer to the EU as a whole. The EU has in place additional regulation that in some cases is stricter than the Montreal Protocol (EEA 2019b). In the context of environmental standards, the most

[6] As Douglass and Fioletov (2011) point out, the 1964-1980 period is taken as baseline due to the reliability of observations and lack of variations rather than for being considered the sustainable level.

relevant feature is the ban of regulated ODS except in very specific uses (Ozone Secretariat United Nations Environment Programme 2009), which can be translated as negative or no consumption of ODS[7].

3.3.2. Regional and local processes

This section addresses the absorption, dispersion and dilution of pollutants in ecosystems. The term pollution can apply to different types of environmental pressure, e.g. air pollutants, plastics, chemicals, etc. Some types of pollution are more relevant in specific ecosystems, but overlaps exist between ecosystems because of how pollutants are transferred to different media. The subsections below describe the main environmental standards used to address the sink functions of terrestrial, freshwater and marine ecosystems.

3.3.2.1. Terrestrial ecosystems

The environmental standards related to terrestrial ecosystems are usually formulated based on the concepts of 'critical levels' and 'critical loads'. Critical levels commonly take the form of concentrations in the atmosphere, while critical loads refer to deposition levels (Cape et al. 2009).

<u>Critical levels</u>

Cumulated exposure to excessive levels of air pollution has a variety of harmful effects on many types of vegetation and impairs their capacity to produce ecosystem services. Such effects include, but are not limited to, changes in yields of biomass or their quality, toxicity, changes in tolerance to stress, etc. (Ashmore and Wilson 1993). The concept of 'critical levels' has been widely used during the last three decades as an environmental standard and in the case of vegetation is defined as "concentration, cumulative exposure or cumulative stomatal flux of atmospheric pollutants above which direct adverse effects on sensitive vegetation may occur according to present knowledge" (Mills et al. 2017, p. 1). Critical levels are commonly adopted for broad vegetation categories such as crops, trees and (semi-)natural vegetation based on the exposure-response relationship of representative receptors. The focus on the most sensitive receptors ensures that the values adopted are generally valid to protect most of the vegetation. UBA (2004) published a manual describing current knowledge around critical levels and critical loads to guide the parties of UNECE Convention on Long-range Transboundary Air Pollution to fulfil their obligations. This knowledge is periodically reviewed based on the latest research.

The adoption of critical levels is a dynamic process informed by empirical studies that report cumulative exposure-response relationships for receptors with varying sensitivities to relevant pollutants. Among those receptors, a representative one is selected for broad vegetation categories such as crops, forest trees and (semi-)natural vegetation, and the critical level beyond which adverse effects appear is agreed between a group of experts. There are three approaches to define critical levels (Cape et al. 2009). In a first stage, quality standards are defined based on the 'no observable effect concentration', i.e. absence of 'measurable difference' from background conditions of the most sensitive

[7] Consumption can be negative when production and imports are lower than exports and destruction of ODS. This usually happens when existing stocks are exported or destructed.

species. When additional evidence is available, critical levels are commonly adopted based on statistical techniques that control for inter-species variation in sensitivity. The use of statistical techniques allows to determine the concentration below which a given percentage of species is protected with a given probability level. The third approach shifts the focus from protecting species to protecting the functioning of the system, which commonly relies on additional evidence on the causal relationships between exposure and changes in ecosystem services. The revision by Mills et al. (2017) provides the most recent update on critical levels for vegetation for four pollutants in Europe: SO_2, NO_X, NH_3 and O_3 (see Table 9).

Critical levels of SO_2, NO_X and NH_3 for European vegetation have remained unaltered since the Egham workshop in 1992 (Ashmore and Wilson 1993). Current SO_2 critical levels distinguish four types of vegetation, with lichens being the most sensitive receptors. The selection is informed by a review in which the effects of SO_2 on vegetation are not harmonised (Bell 1993; WHO 2000). In the case of NO_X, a single critical level for all vegetation was agreed based on growth stimulation and reduction effects reported at low NO_X concentrations for a variety of crops (WHO 2000). For NH_3, critical levels were defined based on 'no-effect' concentrations for lichens and bryophytes (see Table III.4 in Mills et al. (2017)) and for higher plants (Cape et al. 2009). Also, in the case of NO_X and NH_3 there is no homogeneous measure of harmful effects on vegetation.

The knowledge base around O_3 impacts on vegetation has evolved rapidly in the last decades, which has led to critical levels being revised regularly. For some time, cumulative stomatal flux – cumulative O_3 uptake by small pores in leaves – and cumulative exposure have been proposed to characterise critical levels of O_3, while acknowledging that the former is more strongly related to harmful effects on vegetation (Mills et al. 2017). Because the estimation of critical levels based on cumulative stomatal flux requires high volumes of input data into complex models, cumulative exposure is a more straightforward metric to measure exceedance of critical levels at large scales.

The adoption of critical levels for O_3 commonly follows a review of empirical studies for which cumulative exposure-response relationship for relevant receptors are compiled. These commonly include various receptors with different sensitivities to exposure to the pollutant. Among those receptors, a representative one is selected for general categories such as crops, forest trees and (semi-)natural vegetation, and the critical level beyond which adverse effects appear is selected. The choice of a single crop as reference represents a balance between accuracy and data requirements. The quantification of 'adverse effects' can sometimes be done on statistical grounds, e.g. the 5% decrease in yield used as threshold to establish critical levels for crops is justified by Fuhrer et al. (1997) for being the lowest statistically significant change in wheat yield that could be detected at a 99% confidence level. In other cases, the threshold adopted differs from the criterion above. For instance, in the case of trees that statistical criterion would represent a 3% reduction in biomass (Karlsson et al. 2003), yet a 5% reduction is used to determine critical O_3 level. Exposure during the flowering (for crops) or growing season (for trees and other vegetation) to hourly O_3 concentrations above 40 ppb during daylight (accumulated exposure over a threshold; AOT40) is currently used as the metric for critical levels. The excess is expressed as the number of ppm h above the threshold. The selection of the 40 ppb threshold provides the best exposure-response linear fit for wheat yield (Fuhrer et al. 1997) and tree biomass loss (Karlsson et al. 2003).

Critical levels for agricultural and horticultural crops describe the cumulative exceedance that leads to a 5% decrease in the yield of a representative product that is sensitive to O_3 exposure. For agricultural crops, Mills et al. (2007) reviewed the response curve of 19 crops to O_3 exposure and grouped them according to the sensitivity of the response. Wheat was chosen as the representative crop due to its high sensitivity to O_3 exposure and its relevance to European agriculture (Mills et al. 2007). Tomatoes were considered to be representative of horticultural crops on the same grounds (González-Fernández et al. 2014). Based on the studies carried out by Mills et al. (2007) and González-Fernández et al. (2014), the critical levels for agricultural and horticultural crops were set to AOT40 3 ppm h and AOT40 8 ppm h respectively. In the case of tomatoes, quality – monitored as changes in soluble sugar content – were affected at higher concentrations. Karlsson et al. (2003); Karlsson et al. (2007) also reviewed the response of several tree species with diverging degrees of sensitivity to cumulative O_3 exposure. Following these reviews, critical levels for sensitive forest trees were set at AOT40 5 ppm h following a 5% decrease in the biomass of beech and birch (Karlsson et al. 2003; Karlsson et al. 2007). For (semi-) natural vegetation, Mills et al. (2017) report critical levels of AOT40 3 ppm h and AOT40 5 ppm h for vegetation dominated by annuals and perennials respectively for a 10% decrease in above-ground biomass.

Table 9: Selected critical levels of sensitive vegetation in Europe

Receptor	Indicator	Value	Effect	References
Cyanobacterial lichens	SO_2 concentration (annual mean)	10 µg m^{-3}	No occurrence or changes in community structure	Richardson (1988); Ashmore and Wilson (1993); WHO (2000)
Forest trees	SO_2 concentration (Annual mean and Half-year mean (October-March))	20 µg m^{-3}	Reduction in growth and changes in metabolism	Ashmore and Wilson (1993); Bell (1993); Holland et al. (1995); McLeod and Skeffington (1995)
(Semi-)natural vegetation	SO_2 concentration (annual mean and Half-year mean (October-March))	20 µg m^{-3}	Reduction in biomass	Ashmore and Wilson (1993)
Agricultural crops	SO_2 concentration (annual mean and Half-year mean (October-March))	30 µg m^{-3}	No changes in yield reported below critical levels	Ashmore and Wilson (1993); Bell (1993)
All	NO_X concentration (annual mean expressed as NO_2)	30 µg m^{-3}	Growth stimulation and reduction reported around critical levels	WHO (2000)
Lichens and bryophytes	NH_3 concentration (annual mean)	1 µg m^{-3}	Changes in species composition	Mills et al. (2017)
Higher plants	NH_3 concentration (annual mean)	3 µg m^{-3}	Changes in species composition	Cape et al. (2009)
Agricultural crops	AOT40 (measured over 3 months)	3 ppm h	5% decrease in yield	Mills et al. (2007)
Horticultural crops	AOT40 (measured over 3 months)	8 ppm h	5% decrease in yield	González-Fernández et al. (2014)
Forest trees	AOT40 (measured over growing season; 6 months)	5 ppm h	5% decrease in biomass	Karlsson et al. (2003); Karlsson et al. (2007)
(Semi-)natural vegetation: annuals	AOT40 (measured over growing season; 3 months)	3 ppm h	10% decrease in above ground biomass	Ashmore and Davison (1996); Fuhrer et al. (2003)
(Semi-)natural vegetation: perennials	AOT40 (measured over 36months)	5 ppm h	Decrease of 10% in total above-ground or below-ground biomass and/or on the cover of individual species and/or on accelerated senescence of dominant species	UNECE (2006)

Note: The column 'effect' describes the impacts avoided below critical levels. Because in some cases the literature is not clear on the specific impacts avoided, the content of this column should be considered illustrative.
Source: Adapted from Mills et al. (2017)

Critical loads

As opposed to critical levels, which are formulated as concentration, critical loads refer to deposition levels. Thus, critical loads show the deposition levels of pollutants that an ecosystem can remove or buffer without leading to harmful effects (CLRTAP 2017). Specifically, they are defined as "a quantitative estimate of an exposure to one or more pollutants below which significant harmful effects on specified sensitive elements of the environment do not occur according to present knowledge" (Nilsson and Grennfelt 1988).

The state of the art for critical loads is very similar to that for critical levels, in that both are key concepts in the Convention on Long-Range Transboundary Air Pollution (CLRTAP) and, therefore, have evolved in parallel. The first set of critical loads were proposed in Sweden in 1992 (Bobbink et al. 1992) and have been revised periodically since then. The most recent version is provided by CLRTAP (2017), although a European database with information provided by countries is also available (Hettelingh et al. 2017). CLRTAP (2017) describes critical loads of three main pollutants for terrestrial ecosystems: nitrogen-based substances, acidifying substances and heavy metals.

Nitrogen-based substances are particularly relevant because they contribute both to the acidification of soils and to eutrophication, which leads to changes in plant species composition (EEA 2015). Critical loads of nitrogen have been updated through several iterations, mainly based on experimental studies (Bobbink et al. 2015). For illustrative purposes, Table 10 shows the critical loads of nitrogen in woodlands, forests and other wooded lands. CLRTAP (2017) also provides this information for other ecosystem types. The scientific base on which these critical loads are based is summarised in Bobbink et al. (2015).

Table 10: Selected critical loads of nitrogen in woodlands, forests and other wooded land in Europe

Receptor	Value [kg N ha^{-1} yr^{-1}]	Reliability	Effect
Fagus woodland	10-20	Expert judgment	Changes in ground vegetation and mycorrhiza, nutrient imbalance, changes in soil fauna
Acidophilous *Quercus*-dominated woodland	10-15	Expert judgment	Decrease in mycorrhiza, loss of epiphytic lichens and bryophytes, changes in ground vegetation
Mesotrophic and eutrophic *Quercus* woodland	15-20	Expert judgment	Changes in ground vegetation
Mediterranean evergreen (*Quercus*) woodland	10-20	Expert judgment	Changes in epiphytic lichens
Abies and *Picea* woodland	10-15	Expert judgment	Decreased biomass of fine roots, nutrient imbalance, decrease in mycorrhiza, changed soil fauna
Pinus sylvestris woodland south of the taiga	5-15	Quite reliable	Changes in ground vegetation and mycorrhiza, nutrient imbalances, increased N2O and NO emissions
Pinus nigra woodland	15	Expert judgment	Ammonium accumulation
Mediterranean *Pinus* woodland	3-15	Expert judgment	Reduction in fine-root biomass, shift in lichen community
Spruce taiga woodland	5-10	Reliable	Changes in ground vegetation, decrease in mycorrhiza, increase in free-living algae
Pine taiga woodland	5-10	Quite reliable	Changes in ground vegetation and in mycorrhiza, increase occurrence of free-living algae
Mixed taiga woodland with *Betula*	5-8	Expert judgment	Increased algal cover
Mixed *Abies-Picea Fagus* woodland	10-20	Expert judgment	-

Source: CLRTAP (2017)

Critical loads for acidification are derived based on the chemistry and mineralogy of soils (CLRTAP 2017). They depend on the amount of acidity that could be neutralised by the base cations produced by mineral weathering, but are also affected by other factors such as precipitation, vegetation, slope, soil texture, etc. Critical loads in terrestrial ecosystems are commonly defined as a function of the critical values of base cations-to-aluminium and pH. The former responds to the link between increased aluminium concentrations in the soil solution and adverse effects to roots and growth of trees (WHO 2000). Commonly, a base cation-to-aluminium ratio of one is considered, although for some species this can lead to growth reductions above 20% (CLRTAP 2017).

Critical loads of heavy metals in terrestrial ecosystems, on the other hand, can be calculated based on their effects on human health and ecosystem function (Hettelingh et al. 2015), only the latter of which is considered under sink functions. Table 11 shows the types of ecosystems and the rationale for critical loads of heavy metals. In the case of Pb and Cd critical loads are defined based on toxicity of plants, invertebrates and soil microorganisms, while in the case of Hg, critical loads are consistent with the limit proposed by (Meili et al. 2003), who argued that reduced respiration in forest soils occurs at Hg concentrations of 0.5 mg per kg of organic matter content.

Table 11: Critical loads of heavy metals

Receptor	Heavy metal	Effect
Non-agricultural land, arable land, grassland	Pb, Cd	Free metal ion concentration in soil solution in view of effects on soil micro-organisms, plants and invertebrates
Forest	Hg	Total metal concentration in humus layer in view of effects on soil micro-organisms and invertebrates

Source: Hettelingh et al. (2015)

3.3.2.2. Freshwater ecosystems

There is a myriad of pollutants such as pesticides, heavy metals and nutrients that impair the functioning of freshwater bodies, which include surface water and groundwater bodies.

For freshwater, environmental standards for the most important pollutants in the European context have been proposed in the Water Framework Directive (European Parliament and European Council 2000). Following the Directive, the chemical status of surface water bodies is assessed based on compliance with environmental standards defined for substances considered to represent a significant risk to or via the aquatic environment. Environmental standards are commonly expressed as annual average concentrations or maximum allowable concentrations. The former considers chronic effects, while the latter is based on acute toxicity effects. Environmental standards for 45 water pollutants are proposed by experts based on a review of the scientific literature that considers the ecotoxicity and the human toxicity of each substance[8]. For illustrative purposes, a few are shown in Table 12.

[8] Substance-specific background documents can be found here: https://circabc.europa.eu/w/browse/b55f4c81-d664-43db-8b27-264b26a7424b

Table 12: Reference values for selected substances in surface waters

Type	Substance	Annual average concentration (µg/l)	Maximum allowable concentration (µg/l)
Herbicide	Alachlor	0.3	0.3
Polyaromatic hydrocarbons	Anthracene	0.1	0.4
Insecticides	Dichlorodiphenyltrichloroethane, DDT total	0.025	na
Chlorinated solvents	Carbon-tetrachloride	12	na
Metals	Lead and its compounds	7.2	na

Notes: na: not applicable. Where maximum allowable concentration values are marked as 'na', the annual average concentration values are considered protective against acute toxicity effects.

Source: European Parliament and Council of the European Union (2013)

In the case of groundwater, the Directive only sets environmental standards for nitrates and pesticides. The limit for the former is 50 mg/l, while individual pesticides is 0.1 µg/l (0.5 µg/l in total). The decision to cover additional substances is left to Member States.

3.3.2.3. Marine ecosystems

As in the case of freshwater ecosystems, there are many pollutants that have negative impacts on marine ecosystems. Tornero and Hanke (2017) compiled a list of 2,700 substances that could be potentially relevant for marine areas. Thus, environmental standards for marine ecosystems should ensure that their functions are not impaired by excessive pollution.

In Europe, the Marine Strategy Framework Directive contains different descriptors intended to characterise the chemical status of marine waters (European Parliament and European Council 2008a). Three of these descriptors are linked to pollution, namely those referring to eutrophication, other contaminants and plastics. The descriptors intended to characterise the pollution of marine waters are quite vague and are not formulated as environmental standards (EC 2014a). Nonetheless, for pollutant concentrations in water, Tornero et al. (2019) recommends adopting the same reference values as those in the Water Framework Directive, whose scientific rationale has been discussed above.

With regard to plastics, there are no agreed environmental standards related to the plastic load oceans can take without leading to Earth System level changes (Villarrubia-Gómez et al. 2018), but it is safe to assume that the current plastic concentration in oceans is well beyond what could be considered acceptable. Recently, Van Loon et al. (2020) have proposed a reference value of 20 litter items per 100 m beach length for coastal areas. The value represents the 15th percentile of the total litter abundance in European beaches. The authors argue that, in the absence of adequate dose-response data, the reference value reduces the associated negative impact to a sufficiently precautionary level.

3.4. Life support functions

Environmental standards related to life-support functions have mainly been formulated focusing on biodiversity and ecosystems. Biodiversity has multiple dimensions (e.g.

genetic, species, ecosystems), but species have so far received most of the attention in the context of environmental standards. Environmental standards for ecosystems, on the other hand, take the form of indicators of minimum extent or condition. Condition is commonly characterised through pressures or through biological, chemical and physical parameters of state.

3.4.1. Biodiversity

Biodiversity is positively correlated with many ecosystem functions (Cardinale et al. 2012; van der Plas 2019). This relationship tends to be nonlinear and saturating, so that biodiversity loss has relatively small impacts on ecosystem functioning at first, but the latter show accelerating declines with growing biodiversity loss rates (Cardinale et al. 2012). Several attempts have been made to define acceptable biodiversity levels that prevent nonlinear dynamics from taking place.

Rockström et al. (2009b) proposed to use global species extinction rates as proxy of the regulating role of biodiversity because of the potential nonlinear and largely irreversible responses associated with past large-scale biodiversity loss processes. Taking a reference value of 1 extinction per million species-years (E/MSY) from the fossil record, the authors proposed a boundary of 10 E/MSY. In a later update, the same value was proposed as a proxy for genetic diversity (Steffen et al. 2015b). Non-linear dynamics in ecosystem functioning as a result of local and regional thresholds are expected (Rockström et al. 2009b), but whether these thresholds can propagate to the global level is still an open issue (Brook et al. 2013; Hughes et al. 2013; Mace et al. 2014). The choice of an absolute over a probabilistic indicator has also been criticised (Samper 2009).

Acceptable biodiversity levels have also been proposed based on the role biodiversity plays in ecosystem functioning. The relationship between biodiversity and ecosystem functioning is commonly expressed through indicators of species, functional and genetic diversity. Functional and genetic diversity have been argued to outperform species diversity in predicting biodiversity-ecosystem-function relationships (Díaz and Cabido 2001; Cadotte et al. 2008; Flynn et al. 2011; Gagic et al. 2015), although others have claimed that there is insufficient evidence to support this as a general statement (Cardinale et al. 2011). In the absence of global data for functional and genetic diversity metrics (Mace et al. 2014; Steffen et al. 2015b), species diversity indicators have been used to define acceptable biodiversity levels.

In this context, species loss beyond 20% has been documented to affect productivity in terrestrial ecosystems as strongly as other drivers (Hooper et al. 2012). Others have argued that 70% of original species richness should be maintained in each ecosystem (Griggs et al. 2013), although no rationale was given. The proposal that has received the most attention so far is that of Steffen et al. (2015b), who proposed as a preliminary standard to retain 90% of species abundance globally and at biome/large ecosystem level with respect to a time when human intervention was negligible. This value is acknowledged to have a large uncertainty range (90% to 30%). Mace et al. (2018) recently suggested that 70% of ecoregions and 100% of biomes should comply with the 90% abundance target by 2050 in order to meet the vision of the Convention on Biological Diversity (CBD 2010, p. 7), which states that by 2050, biodiversity should be "valued, conserved, restored and wisely used, maintaining ecosystem services, sustaining a healthy planet and delivering benefits essential for all people".

3.4.2. Land use and ecosystem extent

Conversion of forest and other ecosystems to agricultural land is one of the main drivers behind biodiversity loss (Maxwell et al. 2016). For this reason, since the first environmental standards associated with ecosystem extent were proposed a decade ago, most have focused on defining a sustainable agricultural land level.

Within the framework of Planetary Boundaries, Rockström et al. (2009b) proposed that less than 15% of the global ice-free land surface should be converted to cropland [9]. The environmental standard is presented as a highly uncertain global aggregate based on considerations of tipping points related to land use conversion in biomes, and effects on carbon storage and biodiversity loss. Although the relevance of the spatial distribution and intensity of land-system change was acknowledged by Rockström et al. (2009b), the standard has commonly been downscaled in a straightforward manner when adapting it to the national level (Nykvist et al. 2013; Cole et al. 2014; Hoff et al. 2014; Dao et al. 2015; Lucas and Wilting 2018).

A different global standard was proposed by Bringezu et al. (2012) driven by biodiversity loss concerns. Building on the work by van Vuuren and Faber (2009), who assumed that total agricultural land expansion would need to at least stabilize by 2020 in order to halt biodiversity loss. Bringezu et al. (2012) proposed to halt the expansion of global cropland into grasslands, savannahs and forests by 2020. Thus, cropland would be allowed to expand to 1.66 billion ha by 2020. Similarly, although more uncertain due to underlying assumptions, they proposed a sustainability standard of 3.07 billion ha for total agricultural land. Bringezu et al. (2012) allocate these values following the 'environmental space' criteria – i.e. on an equal per capita basis –, which yields 0.2 and 0.37 ha per capita of cropland and total agricultural land by 2030 respectively. As pointed out by O'Brien et al. (2015), 0.2 ha per capita of cropland in 2030 would represent around 12.6% of global ice-free land, which is lower than the maximum value suggested by Rockström et al. (2009b).

Dao et al. (2015) also suggested a maximum value for cropland and urban land, in this case based on policy objectives: a stable surface of urban area per capita until 2050 and halving the global deforestation rate by 2050. These considerations lead to a maximum cropland value of 14.55% of global ice-free land.

The main criticism around the global cropland constraints has focused on the scale at which the standard is defined and the implicit prioritisation of some ecosystem services. Regarding the scale, it has been argued that there is no evidence to support a global standard (Brook et al. 2013), yet it is well established that some regional tipping points such as the irreversible conversion of the part of the Amazon into a savannah as a result of deforestation would have global implications in the climate system and in the water cycle (Lawrence and Vandecar 2014). As for the second point, Bass (2009) argued that there is no reason no prioritise biodiversity loss over other services such as food and fodder associated with certain land use changes. This remains an unresolved issue, for it might not be possible to fully reconcile biodiversity conservation at the scale required with food production for around 9 billion people in 2050. A third argument against global cropland

[9] Although replaced by a boundary based on potentially forested areas in a revision of the framework (Steffen et al. 2015b), the maximum cropland value is still used in national studies.

standards was provided by Running (2012), who proposed to adopt a standard based on net primary production (NPP) for land use. NPP has remained almost constant in the last 30 years (53.6 ± 1 Gt C yr^{-1}). In this context, humans already appropriate around 38% of NPP, while 53% is considered non-harvestable (e.g. plant growth in root systems, preserved land, and wilderness areas where no transportation exists for harvesting). This would leave approximately 5 Gt C yr^{-1} (9%) of harvestable NPP (Running 2012). Erb et al. (2012) challenge the use of NPP as a planetary boundary, since total NPP is influenced by human activities leading to land use changes or increasing productivity. Further, they conclude that there is no evidence to support the contention that the higher the human appropriation of net primary production (HANPP), the less sustainable land management is. In defining the link between HANPP and biodiversity, Haberl et al. (2014, p. 380) argue that "a direct test of the claim that HANPP results in species loss due to a reduction of trophic energy flows so far remains elusive (…)". As such, (HA)NPP should be used as a complementary metric, rather than as a global standard.

More recent proposals are based on spatially explicit models that try to optimise agricultural land use according to different environmental criteria. In this context, Heck et al. (2018b) modelled land use changes to minimise the loss of biodiversity and increase carbon storage capacity setting constraints for minimum food production, water use, and biome-specific maximum biodiversity loss factors. In doing so, the authors generated a set of land use scenarios that can be interpreted as feasibility science-based standards. When aggregated over grid cells, the agricultural land use composition in their 'selected' solution leads to 15.18% of cropland (as a percentage of global ice-free land). This scenario also requires a 22.63% reduction in grazing land compared to 2005 levels. Recently, Usubiaga-Liaño et al. (2019) optimised agricultural land uses across world ecoregions to meet minimum biodiversity standards identified in the literature. In global terms, their results give a range of 4.62-11.17% of cropland and 7.86-15.67% of pasture over global ice-free land, depending on assumptions related to the maturity of secondary vegetation and to how changes in cropland and pasture area are prioritised in the optimisation process.

Based on the proposal by Mace et al. (2014) (c.f. section 3.4.1) and considering previous criticism of the approach by Rockström et al. (2009b), the planetary boundary for land use change was revised by Steffen et al. (2015b). The latest approach shifted from biodiversity to climate change regulation concerns by focusing on forest cover rather than agricultural land. Thus, biome-specific percentage forest cover taking as reference the potential area of forested land in the absence of human intervention was selected as indicator in the latest update of the Planetary Boundaries framework. The reference values were set to 85%, 85% and 50% for tropical, boreal and temperate forests respectively. A weighted indicator on forest cover has been chosen for the global level. Policy goals for global deforestation have also been proposed by the European Commission (EC 2008) considering its implications for climate change and biodiversity loss, and in the Sustainable Development Goals (UN 2015a).

Also in the context of forestry, Bringezu et al. (2012) suggested maintaining the forest area in every continent and to avoid the conversion of primary forest into plantations. The numeric analysis they carried out takes the year 2006 as reference, but this point in time seems to be taken based on the data availability rather than being deemed a reference in which forest resources were not subject to excessive pressures from human intervention. Their analysis used Switzerland as an example and considered land area and the volume of forest resources consumed to derive targets. For land use they converted the amount of forest resources consumed at world level into the corresponding land area and split it in per capita terms to allocate it to countries. Nevertheless, there is no evidence that this value can be considered environmentally sustainable. Considering the huge productivity differences between different regions and forest types, an environmental sustainability standard for forest should focus on the renewability of the resource (O'Brien 2015).

3.4.3. Ecosystem condition

As argued before, ecosystem condition can be characterised through pressure or state indicators. In Europe, the resulting environmental standard takes the form of a qualitative descriptor of condition such as 'excellent', 'good' or 'bad' that results from an expert evaluation that considers the site-specific characteristics of the ecosystem under assessment.

In Europe, most of the environmental standards that can be used to assess ecosystem condition have been laid in environmental legislation. The most relevant pieces of legislation applicable to ecosystem condition are the Habitats Directive (European Council 1992), the Water Framework Directive (European Parliament and European Council 2000) and the Marine Strategy Framework Directive (European Parliament and European Council 2008a).

3.4.3.1. Terrestrial ecosystems

The Habitats Directive covers habitats that are in danger of disappearance in their natural range, that have a small natural range or that present outstanding examples of typical characteristics of one or more of the biogeographical regions present in Europe (European Council 1992). Member States are required to report on the condition of relevant habitats, which, in total, represent around a third of the terrestrial area in EU27 and the UK (EEA 2020b).

Ecosystem condition is referred to as conservation status, which reflects the sum of the influences on the habitat that may affect its long-term distribution, abundance and quality. In broad terms, a 'good conservation status' describes a situation in which a habitat type is prospering (in both quality and extension) and with good prospects to continue to do so in the future (Röschel et al. 2020). Specifically, conservation status of a habitat is defined based on range, area, structure and function. Favourable conservation status is achieved when the following conditions are met:

- its natural range and the areas it covers within that range are stable or increasing;
- the specific structure and functions which are necessary for its long-term maintenance exist and are likely to continue to exist for the foreseeable future; and
- the conservation status of its typical species is good.

Implementing the criteria above requires reference values meant to represent the desired state of the habitat considering its range, area, structure and functions including typical species, and future prospects (Louette et al. 2015). In practice, setting these reference values is not straightforward due to a lack of historical data, problems of identification of habitat types and their specific structures and functions and other factors (Mehtälä and Vuorisalo 2007). Thus, a considerable part of the assessments is based on expert opinions and partial surveys (EEA 2020b).

3.4.3.2. Freshwater ecosystems

The Water Framework Directive is the main legislative piece governing the management of freshwater resources and ecosystems in the EU. As part of the implementation of the Directive, Member States are required to assess the ecological status of their freshwater systems.

The ecological status of surface waters (including artificial and heavily modified water bodies) is determined based on biological, physicochemical and hydromorphological criteria. There are no absolute environmental standards applicable across water bodies, so the ecological status is defined based on the extent to which current values deviate from those attributable to undisturbed conditions. In practice, these three aspects do not need to be monitored to assign an ecological status to a water body. While an assessment of biological parameters is always necessary, the hydromorphological assessment is only mandatory to assign high ecological status to a water body. In this vein, a physicochemical assessment is required to designate high or good ecological status (EC 2003). Except for certain chemical substances, there are no hard fixed standards to determine the overall status of water bodies. The Water Framework Directive provides a normative definition of high and good ecological status. Ultimately, the characterisation of water bodies depends on how Member States characterise the undisturbed conditions and on the intercalibration process aimed at ensuring that the high-good and the good-moderate boundaries in all assessment methods for biological quality elements correspond to comparable levels of ecosystem alteration (EC 2005).

There is some overlap between freshwater ecosystems covered by the Water Framework Directive and Natura 2000 sites, but ‘good ecological status’ and ‘favourable conservation status’ are not equivalent (EC 2011b).

3.4.3.3. Marine ecosystems

The Marine Strategy Framework Directive contains 11 descriptors that should be used by Member States to characterise the environmental status of European marine waters (European Parliament and European Council 2008a). These descriptors address biodiversity, non-indigenous species, commercial fish stocks, food webs, eutrophication, pollution, etc. An initial progress report concluded that the characterisation of environmental status needed to be improved significantly to increase the quality and coherence of the environmental assessments, which led to the adoption of criteria to be used to set environmental standards for each of these descriptors (EC 2017). Nonetheless, to date most of the standards proposed for the marine environment are not measurable (EC 2020).

3.5. Human health and welfare functions

3.5.1. Human health

3.5.1.1. Air pollution

Air pollution is one of the most pressing public health issues in Europe, particularly in urban areas. According to EEA (2019e), around 456,000 premature deaths in Europe were attributable to long-term exposure to air pollution in 2016. In Europe, particulate matter (PM), NO_2 and ground-level O_3 are considered the most relevant air pollutants from a human health perspective (EEA 2019e).

The World Health Organization (WHO) is the authoritative body that provides guideline values for around 30 air pollutants (including the three mentioned above) following expert evaluations of the existing knowledge base linking air pollution and impacts on human health (WHO 2000). Guideline values do not eliminate the risk to human health, but reduce it to what is considered an acceptable level (WHO Regional Office for Europe 2013). Table 13 shows selected values proposed for PM, NO_2 and ground-level O_3. The guideline values proposed are periodically revisited after reviewing the latest scientific evidence available.

Table 13: Reference values for air pollution

Indicator	Reference value	Effect
Annual mean concentration of $PM_{2.5}$	10 µg m^{-3}	This is the lowest levels at which total, cardiopulmonary and lung cancer mortality have been shown to increase with more than 95% confidence in response to long-term exposure to $PM_{2.5}$.
Annual mean concentration of PM_{10}	20 µg m^{-3}	
Annual mean concentration of NO_2	40 µg m^{-3}	Respiratory symptoms in infants.
1-hour concentration of NO_2	200 µg m^{-3}	Increase in bronchial responsiveness among asthmatics
8-hour concentration of O_3	100 µg m^{-3}	Estimated 1–2% increase in daily mortality compared to background concentrations of 70 µg m^{-3}

Source: WHO (2005)

Exposure to PM is linked to respiratory, cardiovascular and other forms of illness. These links have been established both for long- and short-term exposure, yet long-term exposure is considered more relevant (WHO 2005). The existing evidence does not suggest the existence of a threshold below which adverse effects on human health exist. Thus, guideline values are not meant to represent full protection from the effects of PM. For PM lower than 2.5 µm in size ($PM_{2.5}$) the long-term guideline value of 10 µg m^{-3} represents the lowest level beyond which mortality was shown to increase compared to background levels. It should nonetheless be noted that there is evidence that suggests that the current value – set in 2005 – might not be restrictive enough (WHO Regional Office for Europe 2013).

For PM smaller than 10 µm, the existing knowledge base does not allow setting a separate guideline value based on the specific health impacts its exposure leads to. Still, because reducing the emissions of $PM_{2.5}$ and PM_{10} is seen as a joint task, guideline values have been proposed based on average $PM_{2.5}$-PM_{10} ratios. A PM_{10} concentration of 20 µg m^{-3} is considered to provide the same protection level as the 10 µg m^{-3} value for $PM_{2.5}$.

NO_2 is commonly released in combination with other pollutants and therefore, it is not straightforward to isolate its effects on human health (WHO Regional Office for Europe 2013). WHO (2000) reports adverse respiratory effects above 50-75 µg m^{-3} annual average NO_2 outdoor concentrations in children and respiratory symptoms in infants at NO_2 concentrations below 40 µg m^{-3}. In the short-term, 1-hour concentrations above 200 μg m^{-3} have shown an increase in bronchial responsiveness among asthmatics. The additional evidence accumulated since the last revision in 2005 suggests that the guideline values proposed by WHO need to be revisited (WHO Regional Office for Europe 2013).

Exposure to ground-level O_3 is associated with breathing problems, asthma, reduced lung function and lung diseases. Because adverse health effects are visible at concentrations close to background levels, it is difficult to set guideline values at the no-observed-adverse-effect-level or the lowest-observed-adverse-effect level (WHO 2000). So far there is inconsistent evidence on the existence of a threshold above which adverse health effects are detected (WHO Regional Office for Europe 2013). WHO (2005) proposed a guideline value of 100 μg m^{-3} over an 8-hour period, which would increase daily mortality between 1 and 2% compared to the background concentration, clarifying that its exceedance can be occasionally associated with natural factors. This was a downward revision of the previous 8-hour 120 μg m^{-3} concentration, which was connected to a 5% decrease in the pulmonary function of sensitive populations (WHO 2000).

3.5.1.2. Drinking water pollution

As in the case of air pollution, WHO (2011) provides science-based standards for a variety of microbial, chemical, radiological and acceptability aspects of drinking water. Environmental standards for microbes and chemicals are based on the health impacts their intake leads to. For microbial aspects, environmental standards are based on a predetermined tolerable burden of disease. In the case of chemical substances, standards are defined differently for chemicals that require a minimum exposure to have adverse health effects (threshold chemicals) and those for which health impacts have been documented at any given exposure (non-threshold chemicals). For threshold chemicals, tolerable daily intakes are defined considering the level at which adverse effects are apparent. Tolerable daily intakes are then used to set environmental standards that take the form of concentrations in drinking water. For non-threshold chemicals, environmental standards take the form of concentrations in drinking water associated with an estimated upper-bound excess lifetime cancer risk of one additional case of cancer per 100,000 people assuming ingestion for 70 years.

In Europe, the environmental standards proposed by WHO were the starting point of the ones adopted in the Drinking Water Directive (European Council 1998). The Directive, which covers 48 microbial, chemical and indicator parameters that are relevant in the European context, also included environmental standards for a set of pesticides and their degradation products and adopted a more precautionary cancer risk in the case of non-threshold chemicals (EC 2018). Following a recent review by WHO Regional Office for Europe (2017), a proposal for a new drinking water directive was approved (EC 2018). This new proposal includes environmental standards for additional microbial and chemical parameters.

3.5.2. Other welfare

This category covers a wide range of immaterial and often intangible functions of natural capital. Broadly speaking, it includes functions such as recreation, amenity, education, and heritage that are hard to capture through quantitative indicators, but that are nonetheless relevant contributors to welfare. For this reason, in this section only focus on three aspects that can partially be described through indicators (sites of natural relevance, bathing waters and green spaces) are considered, acknowledging that other relevant aspects are not captured.

3.5.2.1. Sites of natural relevance

There are two main inventories of natural sites of special natural and cultural relevance in Europe: the Natura 2000 network and the World Heritage List. The former is linked to the Habitats Directive (European Council 1992), while the latter is part of the World Heritage Convention (UNESCO 1972). In some cases, the sites overlap (EC 2019a). Arguably, Natura 2000 sites are multifunctional in that they provide many ecosystem services linked to the source, sink, life support and welfare functions (EC 2013). Although natural World Heritage sites are also multifunctional, they have been specifically selected for the cultural and heritage values. In this literature review, the environmental standards of the habitats addressed by the Habitats Directive have been described under the life support functions in section 3.4.

The World Heritage Convention covers sites of cultural and/or natural importance. In the context of environmental sustainability only those classified as 'natural' and 'mixed' (i.e. of natural and cultural importance at the same time) are considered. The conservation outlook of relevant sites is regularly assessed focusing on whether the natural and cultural values for which the site was selected are maintained. This is done through desktop research that considers the current state and trend of values, the threats affecting those values, and the effectiveness of protection and management (Osipova et al. 2014).

3.5.2.2. Bathing waters

Exposure to bacteria present in faecal matter, free-living organisms, algae, cyanobacteria and other agents in recreational waters is associated with adverse health outcomes (WHO 2003). In Europe, concentration of pathogens in faecal pollution is the main criterion to characterise the quality of recreational water bodies (European Parliament and European Council 2006), arguably because of the occasional occurrence of episodes related to exposure to agents not related to faecal pollution (Scientific Committee on Toxicity 2001) and because the scientific evidence available does not allow setting guideline values for most of these (WHO 2003).

Enteric illness is the most common negative effect resulting from repeated exposure to faecal pollution, although links to respiratory illnesses have also been found (WHO 2003). WHO (2003) provides guideline values for the concentration of intestinal enterococci in marine waters based on the risk of negative health effects (Table 14). These values were validated in a subsequent review (WHO 2018b), although the validity of the standard is ultimately determined by subjective choices around what is a tolerable health risk. The 2003 report did not find enough evidence to set a similar guidance value for freshwater systems, where the risk of negative health outcomes under the same bacteria concentration is lower compared to marine systems.

Table 14: Reference values for the recreational use of marine water bodies

Body	Indicator	Reference value	Effect
Marine	Concentration of intestinal enterococci	≤40 cfu/100 ml	<1% of gastrointestinal illness risk after repeated exposure <0.3% of acute febrile respiratory illness risk after repeated exposure
		41-200 cfu/100 ml	1-5% of gastrointestinal illness risk after repeated exposure 0.3-1.9% of acute febrile respiratory illness risk after repeated exposure
		201-500 cfu/100 ml	5-10% of gastrointestinal illness risk after repeated exposure 1.9-3.9% of acute febrile respiratory illness risk after repeated exposure

Note: cfu stands for 'colony forming units'
Source: WHO (2003)

Scientific Committee on Toxicity (2001) also supports using E. Coli to characterise the quality of recreational waters, yet this only happens in Europe (WHO 2018b). WHO (2018b) has recently reviewed the reference values adopted in European legislation to categorise recreational water bodies and recommended that the current classification be maintained (Table 15). The intestinal enterococci concentrations in the categories excellent and good represent a risk of 3% and 5% for contracting gastroenteritis and 1% and 2.5% for contracting respiratory illnesses after repeated exposure in marine waters (EC 2002). A concentration of E. Coli 2-3 times higher than that of intestinal enterococci would reflect the same risk (Scientific Committee on Toxicity 2001; EC 2002). As argued above, the risk under the same conditions is considered lower in inland waters (WHO 2003), which was used as a justification to set a higher standard. This assumption is nonetheless refuted by Kay and Fawell (2007).

Table 15: Reference values for the recreational use of inland, coastal and transitional water bodies

Body	Indicator	Reference value	Effect
Inland waters	Concentration of intestinal enterococci	<200 cfu/100 ml	3% of gastrointestinal illness risk after repeated exposure 1% of acute febrile respiratory illness risk after repeated exposure
		201-400 cfu/100 ml	5% of gastrointestinal illness risk after repeated exposure 2.5% of acute febrile respiratory illness risk after repeated exposure
	Concentration of Escherichia coli	<500 cfu/100 ml	3% of gastrointestinal illness risk after repeated exposure 1% of acute febrile respiratory illness risk after repeated exposure
		501-1000 cfu/100 ml	5% of gastrointestinal illness risk after repeated exposure 2.5% of acute febrile respiratory illness risk after repeated exposure
Coastal and transitional waters	Concentration of intestinal enterococci	<100 cfu/100 ml	3% of gastrointestinal illness risk after repeated exposure 1% of acute febrile respiratory illness risk after repeated exposure
		101-200 cfu/100 ml	5% of gastrointestinal illness risk after repeated exposure 2.5% of acute febrile respiratory illness risk after repeated exposure
	Concentration of Escherichia coli	<250 cfu/100 ml	3% of gastrointestinal illness risk after repeated exposure 1% of acute febrile respiratory illness risk after repeated exposure
		251-500 cfu/100 ml	5% of gastrointestinal illness risk after repeated exposure 2.5% of acute febrile respiratory illness risk after repeated exposure

Note: cfu stands for 'colony forming units'
Source: EC (2002)

3.5.2.3. Green spaces

Research suggests that exposure to green spaces positively contributes to physical and mental health (Bell et al. 2014). Thus, improving access to green spaces has become a relevant goal of urban planning, especially in times of COVID (Geary et al. 2021).

There are multiple elements that make it difficult to quantify access. From inconsistent definitions of green spaces (Taylor and Hochuli 2017) to the operationalisation of access in quantitative terms (Woldeamanuel et al. 2020), which includes, for example, metrics of distance and travel time. The latter is considered to be more suitable (Jalkanen et al. 2020). Nonetheless, an agreement around acceptable time travel times is lacking. Several authors use the 10-minute walking figure (Poelman 2018; Woldeamanuel et al. 2020), although other figures exist (Kabisch et al. 2016).

3.6. Discussion

The previous section offers a long overview of reference values proposed for a wide range of environmental and resource issues. In general, the literature in which the previous section is based is very scattered, which shows that most environmental standards have been proposed with a specific topic in mind, rather than as part of a holistic environmental sustainability vision that encompasses broader environmental and resource aspects. Exceptions could be the environmental standards proposed as part of the Planetary Boundaries framework (Steffen et al. 2015b), which consider processes related to the climate system, biosphere integrity, freshwater, biogeochemical cycles, etc. The following subsections discuss the literature using two lenses. First, the overall adequacy of individual reference values is discussed. Second, the literature reviewed is contextualised looking at existing environmental standards from an integrated perspective in which more general features and limitations to their use are discussed. All this is done within a European perspective given the geographical scope of this thesis.

3.6.1. Overview of environmental standards

Source functions

Source functions can be split into renewable and non-renewable resources. In general, the environmental standards for resources take the form of exploitation rates that are deemed environmentally sustainable.

In the case of renewable resources, sustainable exploitation rates are based on the regenerative capacity of the resource. This is the case, for instance, for forest utilization rates (EEA 2017), water exploitation rates for surface water and groundwater (Raskin et al. 1997; EC 2009) or concepts such as maximum sustainable yield for fish (Meltzer 2009). In the case of fish, the main methods used in Europe to define overexploitation are based on criteria on stock abundance, population age and size distribution, and reproductive capacity, although the specific standards and reference values can differ (EC 2010; FAO 2011).

For non-renewable resources, scarcity is key and thereby sustainable exploitation rates are defined considering the exploitation potential of the resources over a given timeframe. For soils, tolerable soil erosion rates based on the formation rate of soils are used as environmental standards (Verheijen et al. 2009). Other factors such as the content of organic matter, salinization and sealing are also linked to the functioning of soils, but lack a credible environmental standard (Loveland and Webb 2003; Huber et al. 2008). Land degradation neutrality, which measures non-declining carbon stocks, land cover and land productivity compared to 2015, has been adopted as a goal in the SDGs (IAEG-SDGs 2016), but without a baseline that can be considered sustainable, this cannot be considered a science-based standard. Regarding the extraction of abiotic raw materials such as metal ores, non-metallic minerals and fossil energy carriers, the environmental standard could take the form of use-to-availability ratios (with different possibilities to represent availability) that indicates the time the extraction of a given material could be sustained under projected extraction rates. To date, only Henckens et al. (2014) has ventured to propose what an adequate timeframe could be, but as the authors acknowledge, this is arbitrary. Likewise, several resource types exist for which, rather than scarcity, the environmental impacts arising from its use represent the main limitation factor. Examples include the extraction of fossil fuels (McGlade and Ekins 2015) and metals (Desing et al. 2020). Reference values related to the consumption of raw materials also exist (Schmidt-Bleek 1993; Bringezu 2009, 2011, 2015), but consumption of raw materials is commonly used as a proxy for environmental pressures (Steinmann et al. 2017) and is therefore not representative of the source functions of natural capital.

Independent from the renewability of the resource, sustainable exploitation rates can be defined at different scales from global to local. Thus, the adequacy of the geographical scale needs to be assessed on a case-by-case basis. For instance, for metals and fossil fuels, sustainable extraction rates can be defined at deposit, region, country or global level, but given that these are commodities that are traded globally, the latter seems to be more meaningful. Conversely, for fish resources, focusing on stocks rather than the total population of a species is more reasonable. Freshwater is a good example of a resource for which the geographical scope of standards is changing over time. Originally, the sustainable exploitation rate was defined based on annual freshwater availability at the national level (Raskin et al. 1997) and in some cases still is (IAEG-SDGs 2021). In Europe, the standard is now defined at river basin level and instead of considering annual resources, it integrates the temporal variability dimension by focusing on quarterly water flows instead (Faergemann 2012).

All in all, in Europe the use of environmental standards related to renewable resources seems to be widespread when assessing the status of forest (EEA 2017), fish (EEA 2019d) and freshwater resources (EEA 2018b, 2018c). For other topics such as food resources, which could be characterised through pollination or soil productivity, no environmental standards have been found. In the case of non-renewable resources, the use of reference values is mostly restricted to soil erosion (Panagos et al. 2020). In the case of other non-renewable resources, proposed reference values are arbitrary (Henckens et al. 2014).

Sink functions

Sink functions refer to the capacity of natural capital to absorb, disperse or dilute wastes to reduce potential harms. They are split in two main groups. The first one addresses emissions affecting global processes, while the second addresses waste flows that lead to regional or local environmental degradation. In the case of global processes such as global warming and the depletion of the ozone layer, environmental standards can take the form of changes to mean global temperature increases (Schellnhuber et al. 2016) and thickness of the ozone layer respectively (Rockström et al. 2009b). However, to be applicable at the national level, these global standards need to be translated to country emissions of GHG and ODS. Given past and current trends, it seems reasonable to state that country emissions of GHG and consumption of ODS will eventually have to fall to near zero, or even negative values as is already the case for the latter in Europe (EEA 2019a). At the global level, the consumption of ODS has already decreased more than 99% compared to the mid-1980s (Ozone Secretariat United Nations Environment Programme 2019). Because of this, the implementation of the Montreal Protocol and its amendments represent one of the most evident success stories of global environmental policy to date.

The case of GHG emissions is somewhat different. Many countries have formulated targets to get to net zero emissions – in most cases by 2050 or later - (UNFCCC 2021). This implies that countries will have emission levels above the targets for some decades at least. From an environmental sustainability perspective, an environmental standard needs to represent emission levels that can be sustained over time, which will not be the case until countries reach or are close to reaching net zero emissions. Providing specific figures for sustainable near zero emission levels is not straightforward. Different approaches exist to do so (Höhne et al. 2014), and they lead to different results (van den Berg et al. 2019). Given that downscaling principles have great political implications, it is unlikely that an agreement will be reached around the method to allocate responsibilities. In the absence of such a method, simplicity and transparency could help in the context of this thesis. For instance, downscaling the carbon budgets in Table 8 allocating emissions on an equal-per-capita basis using cumulative population figures leads to 0.5 and 2.5 tonnes CO_2 per capita to meet the 1.5ºC target with a 67% probability (420 Gt CO_2 globally) and the 2ºC target with a 33% probability (1,170 Gt CO_2 globally) respectively. These could be used as environmental standards for CO_2 emissions in the following chapters.

The second group of environmental standards in the sink function addresses waste flows that lead to regional or local environmental degradation. Because the effects of pollutants at these scales depend on the characteristics of the receptors, environmental standards take different forms and tend to be location-specific, which, in this case, results in a very Europe-centric set of standards, many of which have been established as part of environmental policies or legislation. In terrestrial ecosystems, they are often represented through critical levels and critical loads. The former refer to pollutant concentrations in the air, while the latter refer to the deposition of pollutants on land and vegetation. In freshwater and marine systems, environmental standards usually take the form of pollutant concentrations in waters, which vary depending on whether the focus is set on short-term or chronic effects of pollution. When data is available, these standards are based on the negative impacts they have on ecosystems and (sometimes) on humans. In practice, environmental standards have only been proposed for a small fraction of all available substances (Brack et al. 2018), so assessing the chemical status of ecosystems requires shortlisting the pollutants that are most important. There are also differences when using the environmental standards in assessments. For instance, in terrestrial ecosystems, transgression of standards has been assessed for individual substances separately (Fagerli et al. 2020; Horálek et al. 2020). In freshwater systems, on the other hand, individual substances have been considered, but chemical status takes the form of a composite indicator that uses the 'one out, all out' rule. In other words, for the chemical status of a freshwater system to be considered good, it requires that system to comply with the environmental standards of each individual substance (EEA 2018b). Ideally, standards should not be restricted to the effects of individual pollutants, but also consider the effects they have when combined.

Life support functions

Environmental standards of life support functions are intended to depict the status of the elements of natural capital that underpin life on Earth. Standards have been proposed for the status of biodiversity, since this is an important predictor of the functioning and stability of ecosystems (Cardinale et al. 2012), and for the extent and condition of ecosystems.

For the status of biodiversity, proposed standards have taken the form of global species extinction rates and species abundance (Steffen et al. 2015b), although other aspects of biodiversity for which standards are not available have been identified (Mace et al. 2014). To date, the most well-known biodiversity standards are generic and therefore embed significant uncertainties because they have been formulated without capturing the specific functions they fulfil in their respective ecosystems. Standards of ecosystem extent, which take the form of limits to agricultural land (Rockström et al. 2009b; Bringezu et al. 2012; Usubiaga-Liaño et al. 2019), share a similar problem, since they have been formulated with biodiversity conservation as a central goal.

Ecosystem condition standards tend to be more complex than the biodiversity and ecosystem standards reviewed. The latter are formulated as single indicators, while ecosystem condition is commonly assessed against multiple criteria, some of which relate to biodiversity. Thus, the ecosystem condition standards reviewed embed biodiversity considerations.

Ecosystem extent and condition metrics have been recently integrated in the Ecosystem Accounts section of the System of Environmental Economic Accounting (UNDESA 2021), which intends to provide a harmonised framework for ecosystem accounting. In the case of ecosystem condition, relevant indicators are related to reference conditions, which can simply show ecosystem condition at the starting point of the accounting exercise, or represent a state that is relatively undisturbed or undegraded by humans, or a situation in which the ecosystem is in relative stability. Despite the conceptual guidance on ecosystem extent and condition accounting, the System of Environmental Economic Accounting does not provide specific environmental standards. A review by Maes et al. (2020) showed that the use of reference conditions differs between countries. Thus, the System of Environmental Economic Accounting is still far from being implemented in a consistent manner across countries.

In Europe, ecosystem condition standards are defined by the relevant legislation. For terrestrial ecosystems, parameters on range, area, structure and function are used to define good quality (Röschel et al. 2020). The condition of freshwater and marine ecosystems, on the other hand, is determined based on a wide range of biological, physicochemical and other parameters (EC 2003, 2017). Because of the number of parameters considered in the definition of ecosystem condition, assessments use composite indicators that assign a qualitative score (e.g. good, bad) to individual ecosystems (EEA 2018b, 2020b). Currently ecosystem condition standards are more developed than biodiversity status and ecosystem extent standards in Europe.

Beyond the biodiversity and ecosystem standards reviewed, several targets to increase the extent of protected areas have been formulated (CBD 2010). Arguably, protected areas have the goal to protect biodiversity and to improve ecosystem health, but they are not synonymous of good biodiversity status and good ecosystem health (Jones et al. 2018; Wolf et al. 2021) and therefore cannot be used as proxies of environmental standards of life support functions.

<u>Human health and welfare functions</u>

The functions in this group are linked to human health and other aspects of well-being such as recreation, culture, spirituality, etc. In relation to human health, environmental standards are formulated as maximum concentrations of air pollutants in indoor and outdoor environments (WHO 2005), in drinking water (European Council 1998) or bathing sites (EC 2002). All these standards are based on the health impacts of different pollutants on humans, although bathing sites are also associated with recreation. It is particularly striking that air pollution targets in public policy are often weaker than environmental standards considering that the latter have been proposed by a well-established institution such as the World Health Organisation (Kutlar Joss et al. 2017; ECA 2018). While pollution standards take the form of concentrations, many assessments either complement or even replace exposure indicators by mortality and disability indicators. The latter indicate (premature) deaths and disability-adjusted life years respectively. To date, there are no standards for these aspects of pollution, but several researchers have interpreted the SDGs as requiring zero mortality (OECD 2019; Sachs et al. 2020), which would be aligned with the maintenance of the health function. Nevertheless, concentration-based environmental standards are more established and rely on an extensive literature base.

In the case of other welfare functions, standards are lacking for most of the non-use values of natural capital. While access to green areas (Poelman 2018), bathing areas and the condition of natural and mixed World Heritage sites (Osipova et al. 2014) cover some recreational and aesthetic values of natural capital, they are just a subset of the elements of natural capital that have aesthetic, spiritual, religious, historic, scientific, educational information, cultural and artistic value.

3.6.2. General features of environmental standards

The main goal of the literature review is to find environmental standards that can be used in the next chapter to compute two environmental sustainability metrics of the ESGAP framework. To that end, the previous lines are intended to facilitate this task by providing insights that help assess the adequacy of the environmental standards proposed in the literature. Beyond individual standards, it is also relevant to look at them altogether in order to be able to interpret the results critically.

Environmental standards take different forms depending on the function they address. For instance, standards of source functions describe exploitation rates to assess the renewability or the scarcity of resources, while standards of life support functions, for instance, the condition of biodiversity or ecosystems. This also holds true for the use of individual or composite indicators that consider multiple standards depending on the function to be assessed. Because of the differences between the broad function categories and within them (e.g. global vs local processes), environmental standards do not have a homogeneous meaning in that they can refer to acceptable health risks, acceptable environmental impacts, precautionary expert guesses, or judgements about safe distance from tipping points. In this context, the level of consensus around standards differs considerably. In all cases though, their transgression flags a potential problem that requires further policy attention.

In Europe, many environmental standards are based on environmental policy and legislation. This does not mean that environmental policy targets are aligned with environmental standards as has been argued before (Kutlar Joss et al. 2017; Doherty et al. 2018; UNEP 2020). Nonetheless, the need to monitor the state of different elements of natural capital has led to the development of specific criteria in some areas. Examples include the chemical and environmental status of terrestrial, marine and freshwater ecosystems (European Council 1992; European Parliament and European Council 2000, 2008a) or the adoption of the Convention on Long Range Transboundary Air Pollution as a result of the acidification of Scandinavian forests in the early 1970s, which led to the development of the concept of critical loads (UNECE 2015). Because of this, many of the environmental standards reviewed are specific to Europe. This needs to be taken into account when implementing the ESGAP framework in non-European countries, since the standards and, consequently, the indicators that can be used to characterise environmental sustainability will differ between countries.

Something else that needs due consideration is the adequate coverage of environmental and resource topics by the standards reviewed. No standards have been found for food resources, some aspects of soil resources, and many welfare aspects related to natural capital. Likewise, some standards have been found to be less robust than others, e.g. biodiversity status, ecosystem extent, extraction of abiotic resources or plastic pollution. Consequently, not all the relevant topics are covered with the same degree of scientific rigour. The existence of environmental standards does not indicate that some elements of natural capital are more relevant than others. Rather, it shows that a suitable reference value is lacking to judge its environmental sustainability. As the knowledge base improves, existing environmental standards might change, or new ones might be formulated. This is to be kept in mind for a future update of this indicator framework.

3.7. Conclusions

This chapter provides an overview of a wide variety of reference values proposed for different environmental and resource topics, alongside their rationale. These have either been taken from the scientific literature or from relevant environmental legislation informed by expert input.

The goal of environmental standards is to provide a science-based reference value that allows assessing whether a given function of natural capital can be sustained over time. Thus, it helps contextualise the information provided by natural capital indicators. Nonetheless, not all the aspects of natural capital have environmental standards. In order to gain a much more detailed picture of the status of specific elements of natural capital or to monitor the state of the environment, a set of more comprehensive indicators – many of which lack environmental standards – is needed, such as those proposed in the Natural Capital Indicator Framework (Fairbrass et al. 2020a) or state of the environment reports (e.g. EEA (2019c); UN Environment (2019a)).

In the following chapters, the overview presented here is used to inform the indicator selection process in SESI and SESPI. To that end, the suitability of the environmental standards presented here are considered on a case-by-case basis.

4. Strong Environmental Sustainability Index

4.1. Introduction

Previous chapters have highlighted the need for better metrics to monitor the environmental sustainability of nations. For such metrics to be aligned with the strong sustainability proposition, they need to reflect whether the long-term maintenance of the diverse functions of natural capital is threatened. The latter requires comparing the current situation with reference values that represent sustainability conditions. In chapter 2, these reference values are referred to as environmental standards. Environmental standards have been reviewed in chapter 3.

The ESGAP framework proposes two indices to monitor the environmental sustainability performance of countries and progress towards it. The use of indices has both benefits and drawbacks as summarised in Table 16. On the positive side, they can summarise complex and multidimensional concepts by showing the big picture through a single metric that captures the attention of relevant audiences such as politicians and the general public. Nonetheless, the use of indices can lead to suboptimal or even poor decisions if interpreted in isolation or when choices made during their construction are not based on sound principles.

Table 16: Pros and cons of indices

Benefits	Drawbacks
• They can be used to summarise complex or multidimensional issues, in view of supporting decision-makers. • They provide the big picture. They can be easier to interpret than trying to find a trend in many separate indicators. They facilitate the task of ranking countries on complex issues. • They can help attract public interest by providing a summary figure with which to compare the performance across countries and their progress over time. • They could help to reduce the size of a list of indicators or to include more information within the existing size limit.	• The simple "big picture" shown may invite politicians to draw simplistic conclusions. Thus, indices should be used in combination with the sub-indicators to draw sophisticated policy conclusions. • They may send misleading, non-robust policy messages if they are poorly constructed or misinterpreted. • Their construction involves stages where judgements must be made. These judgements should be transparent and based on sound conceptual and statistical principles. • The selection of indicators and weights could be the target of political challenge.

Source: Adapted from Saisana et al. (2005)

SESI, which is described and computed in this chapter, characterises the environmental sustainability performance of a country at a given point in time. Thus, it sheds light on the first research question identified in the introduction: *Are European countries environmentally sustainable?*

Like other indices, SESPI uses a series of indicators that are then normalised, weighted and aggregated across different levels in order to generate a final score. These steps also apply to SESPI, described in detail in chapter 5. The process of building an index is not straightforward. The OECD and JRC (2008) published the most comprehensive handbook to date on how to construct composite indicators. The manual describes ten steps that cover the different stages of the process: from the development of the theoretical framework to the visualisation of the results. An updated version is being prepared with a reorganised process (Table 17).

Table 17: Steps recommended by the JRC to construct an index

Step	Name	Description
1	Theoretical framework	The theoretical framework provides the basis for the selection and combination of variables into a meaningful index that is fit for purpose.
2	Indicator selection	The selection of data and indicators should be based on the analytical soundness, measurability, country coverage, and relevance of the indicators to the phenomenon being measured and their relationship to each other.
3	Data treatment	After assembling a set of indicators, missing data can be imputed, outliers treated, and transformations can be applied to indicators where necessary and appropriate.
4	Normalisation	Normalisation brings indicators onto a common scale, which renders the variables comparable.
5	Weighting	When indicators are aggregated into a composite measure, they can be assigned individual weights. This allows the effect or importance of each indicator to be adjusted according to the concept being measured.
6	Aggregation	Aggregation combines the values of a set of indicators into a single summary 'composite' or 'aggregate' measure.
7	Statistical and conceptual coherence analysis	This can be used to study the overall structure of the dataset, assess its suitability and coherence, and assist in the revision of the choices made in previous steps (e.g. weighting and aggregation).
8	Uncertainty and sensitivity analysis	Uncertainty analysis quantifies the uncertainty in the scores and ranks of the index, as a result of uncertainty in the underlying assumptions. Sensitivity analysis quantifies the uncertainty caused by each individual assumption, which identifies particularly sensitive assumptions which might merit closer consideration.
9	Identify narratives and links to other metrics	Develop relevant narratives and stories to communicate the results. The scores of the index (or its dimensions) should be correlated with other relevant indicators to identify linkages.
10	Visualisation	Indices are ultimately a communication tool, which can be greatly enhanced by proper visualisation.

Source: Adapted from JRC (2019)

The theoretical framework (step 1) has already been described extensively in chapter 2. In this chapter, the remaining steps are addressed explicitly or implicitly. The choices made in relation to indicator selection, data treatment, normalisation, weighting, aggregation, statistical coherence and sensitivity analysis (steps 3-8) and the underlying rationale is presented in separate sections (4.2-4.8). Combined, they represent the methodology of SESI. Steps 9 and 10 are implicitly addressed in the results and discussion sections (4.9 and 4.10). Chapter 6 represents an extension of step 9, which deals with the link between SESI and other relevant metrics.

4.2. Indicator selection

The selection of indicators is a critical step in the construction of an index. This is particularly true in an index of strong sustainability, since metrics of weak sustainability underestimate environmental problems, while the remaining metrics do not always reflect their urgency. In this thesis, the term 'strong environmental sustainability indicators' (SES indicators) is used to refer to the indicators that are normalised, weighted and aggregated in order to generate SESI.

4.2.1. Criteria for selection

Different criteria can be used to select metrics to populate indicator systems or the structure of indices (e.g. Srebotnjak et al. (2009); UNSD (2015); Eurostat (2020a)). Here the criteria used by Eurostat for their 2020 SDG indicator set is used as reference. Eurostat uses three main criteria to select the indicators used to monitor progress towards the SDGs: policy relevance, statistical and methodological soundness, and data quality. These criteria are adapted to the ESGAP framework as follows.

4.2.1.1. Relevance

The theoretical framework is arranged around four broad environmental function categories, each of which needs to be characterised by appropriate SES indicators. For an indicator to be relevant, it needs to have the following three characteristics:

- First, the indicator needs to be linked to the environmental functions of natural capital. In the case of the environmental functions used in the ESGAP framework (source, sink, life support and human health and welfare), it should be an indicator (or proxy) of environmental pressure, state or impact in most cases, except in the case of human health and welfare functions, where social state indicators would be most appropriate.
- Second, an appropriate reference value is required against which performance can be measured. That reference value should be defined through science-based environmental standards that ultimately represent the conditions under which the functioning of natural capital is not altered in a way that it threatens its capacity to provide ecosystem services in the long-term.
- Thirds, the indicator must be relevant at the national level, for this is the geographical scope for which SESI is produced.

There are hundreds of environmental indicators being produced at different spatial scales in Europe as a result of activities related to natural capital accounting, environmental economic accounting and monitoring of environmental policies. Nonetheless, the literature on environmental standards is relatively limited and certainly not advancing at the same pace. As a result, environmental standards are the main limiting factor in selecting SES indicators and these therefore need to be used as a starting point in the selection process.

4.2.1.2. Statistical and methodological soundness

Eurostat (2020a) refers to five key requirements that the SDG indicators should meet:

- Readiness of statistical production: indicators must have at least one data point ready to use and published by their producer.
- Sustainability of statistical production: regular data production must be ensured, preferably by an official mandate and by adequate human (including quality of staff) and financial resources.
- Sound methodology and procedures: indicators and their underlying data must be produced according to a well-founded methodology and procedures.
- Accessibility and transparency: data on indicators must be accessible online and information on their data sources, methods of computation, etc. must be publicly available.

- Compliance: indicators must comply with international or EU standards where such standards exist (agreed methodology, definitions, classifications, standards and recommendations).

As they argue, the indicators provided by official statistical offices and well-established international institutions would meet the above criteria, although exceptions have been documented where official statistics failed to be reliable (Mooney et al. 2021). Only in cases where relevant indicators are missing could these criteria be relaxed.

4.2.1.3. Data quality

When it comes to data quality, Eurostat uses a score-based system across a range of quality criteria such as a frequency of dissemination, timeliness, time coverage, data comparability, etc. Given that this first version of SESI and SESPI is meant to be a proof of concept, the original Eurostat data quality criteria has been relaxed a bit as shown in Table 18. Given that ultimately the thresholds used in each criterion are subjective, these should be revised should the SESI and SESPI be used in official indicator reporting activities.

Table 18: Data quality criteria used for SES indicator selection

Criterion	Rating				Comments
	High (3 points)	**Medium (2 points)**	**Low (1 point)**	**Insufficient (0 points)**	
Frequency of dissemination	1 year	2 years	>2 years or not disseminated regularly, but data can be produced with reasonable effort	Not disseminated regularly	SESI is initially intended to be updated every 1-2 years.
Timeliness (T = base year)	T - 1 year	T - 2-5 years	T - 6-10 years	T - >10 years	For the indices to be relevant the data points should be as recent as possible
Geographical coverage	All 28 countries	80-99% countries (23-27)	67-80% countries (19-22)	<67% countries (<19)	The JRC (2019) recommends a minimum threshold of 2/3 of country coverage.
Geographical comparability	All 28 countries	67-99% countries (23-27)	Limited	-	Rating based on comparability according to the most recent data points. Data that is not geographically comparable across can still be used to calculate SESI for individual countries.
Time coverage	≥3 data points in periods of <5 and 5-15 years	≥2 data points in periods of <5 or 5-15 years	1 data point	Data not available	Only one data point is required to calculate the SESI, but SESPI requires at least two data points. Depending on the period for which data is available, progress in the short- and mid-term can be calculated with SESPI.
Temporal comparability	All years	≥2 data points, but not all	Limited	-	Data that is not comparable across time can be used to calculate SESI, but not SESPI.

Source: adapted from Eurostat (2020a)

4.2.2. Selection process

30 indicators have been shortlisted as potential candidates to build SESI based on the literature on environmental indicators and standards. The candidate indicators have been assigned to one of the environmental functions and sustainability principles shown in Table 4. These have then been used as dimensions in the aggregation process to build SESI. The mapping is based on the author's judgement. It should be noted that there is no perfect fit between the functions and the indicators, since an indicator can be related to more than one function (e.g. climate change, which has pervasive effects across many environmental areas). This initial list builds on the literature review in chapter 3 and considers feedback obtained in different meetings as part of a related project funded by the French Development Agency. Indicators have been assessed against the criteria

described above in a sequential process. This process allowed discarding the indicators that did not meet the minimum relevance, soundness and data quality criteria.

4.2.2.1. Relevance

Table 19 maps the 30 candidate indicators identified to possible environmental standards. The entries of the table are colour coded based on the existence of relevant reference values. Most indicators represent an environmental or social state that shows a percentage of ecosystems, water bodies, population, etc. that meets an environmental standard. When such state indicators were not available or were not relevant at the country level (e.g. when describing global processes), environmental pressure indicators were used. In the table, green shading indicates the existence of a suitable environmental standard. Yellow shading indicates the existence of a reference value that is not deemed to be an environmental standard. Red shading refers to the absence of an environmental standard. The indicators with green shading have been assessed against the soundness criteria in the next subsection.

<u>Indicators for source functions</u>

Indicators for the source function cover renewable and non-renewable resources. Renewable resources include forest, fish, groundwater and freshwater resources. The environmental standards for these indicators tend to describe exploitation rates (e.g. extraction vs annual availability) that are deemed environmentally sustainable (Raskin et al. 1997; EC 2009; EEA 2017), except in the case of fish resources, which represents an exploitation status that uses criteria on fishing mortality and spawning stock biomass to define overexploitation (EC 2010).

Indicators of non-renewable resources are restricted to soil resources, in this case represented through soil erosion. The environmental standard is defined as the tolerable soil erosion rate (Verheijen et al. 2009). Other aspects of soil resources such as the content of organic matter, salinization, sealing and land productivity lack environmental standards (c.f. chapter 3).

Regarding abiotic raw materials, these can be approached from the side of extraction or consumption, both of which can be formulated through indicators used in economy-wide material flow analysis. The environmental standard for the extraction of raw materials could take the form of a reserves-to-production ratio that indicates the time the extraction of a given material could be sustained under projected extraction rates with existing technologies. Such standard would be conceptually aligned with the source function of natural capital and consider scarcity issues. Given the unequal distribution of raw materials, material-specific reserves-to-production ratios should be formulated at the global level and extraction quotas allocated to countries for the extraction of abiotic materials to be further considered in the next steps. Nonetheless, science-based standards are lacking. Furthermore, grouping material categories into the broad abiotic categories used in economy-wide material flow analysis (metal ores, non-metallic minerals and fossil energy materials/carriers) seems particularly problematic in this case. From the consumption perspective, several authors have proposed environmental standards (Schmidt-Bleek 1993; Bringezu 2009, 2011, 2015), but these have been adapted over time without solid arguments. But more importantly, national consumption of raw materials is commonly used as a proxy for environmental pressures (Steinmann et al.

2017) and is therefore not representative of the source functions of natural capital. As a result, indicators of abiotic raw material extraction are excluded from further scrutiny, although the environmental impacts their extraction and use lead to are implicitly considered in other functions.

Indicators for sink functions

Sink functions refer to the capacity of natural capital to absorb wastes and can be split depending on the geographical scope of the processes these wastes disrupt. At the global level, climate change and the depletion of the ozone layer can be addresses through indicators of emission of GHG and consumption of ODS. In both cases, these pressure indicators will have to fall to near zero or negative values. In Table 19, global standards for global processes have been marked with a green shading and downscaled to the national level at a later stage.

The second group of indicators in the sink function is focused on environmental degradation process at the regional and local levels. Terrestrial ecosystems are characterised through indicators of exceedance of critical levels of ozone and critical loads of acidification, eutrophication and heavy metals. These pressures have been selected based on the availability of environmental standards and data, although many others could be added. Ecosystem-specific critical levels and loads of pollutants have been reported in different sources (Karlsson et al. 2003; Karlsson et al. 2007; Mills et al. 2007; Hettelingh et al. 2015; CLRTAP 2017; Hettelingh et al. 2017). Freshwater ecosystem indicators take the form of dichotomous composite metrics that consider compliance (or no compliance) with concentration of pollutants in surface waters and groundwater as defined by the relevant European legislation (European Parliament and European Council 2008b; EC 2009). For marine waters, the European legislation characterises their environmental status based on different descriptors (EC 2017), only some of which (e.g. eutrophication, litter and other contaminants) are related to the sink function. Marine areas beyond national jurisdiction could be considered as a fourth type of ecosystem, but they have been left out due to the difficulty of assigning responsibilities to countries for excessive levels of pollution. This is an area that should be further explored in the future to address relevant environmental problems such as chemical or plastic pollution.

Indicators for life support functions

Indicators for life support functions are intended to reflect whether biodiversity and ecosystem health is maintained. In the initial indicator list, these elements are characterised through metrics on the status of biodiversity in and ecosystem condition for different types of ecosystems.

For terrestrial ecosystems, we use the conservation status of terrestrial ecosystems of European interest as defined in the relevant European legislation (European Council 1992). The habitats considered only cover around one third of the terrestrial area of the EU Member States (EEA 2020b). The Local Biodiversity Intactness Index could be used as a proxy for functional diversity with the environmental standard proposed in the Planetary Boundaries framework (Steffen et al. 2015b), but given that the conservation status of terrestrial ecosystems considers function as one of the criteria to assess condition, the Local Biodiversity Intactness Index is not considered further due to redundancy. For freshwater and marine ecosystems, we use the wide range of biological, physicochemical and other parameters of ecosystem condition defined in the relevant environmental legislation as environmental standards (EC 2003, 2017). The resulting indicators are composite metrics of ecosystem condition.

Life support functions could also be represented by additional indicators of key elements of natural capital. The most obvious example would be climate and average temperature increase in a country compared to pre-industrial levels. Nonetheless, this type of indicator is not responsive to policy interventions in the short- to mid-term and therefore not relevant in the context of the ESGAP framework.

Indicators for human health and welfare functions

The indicators in this category are split into two groups: human health and amenity. The former covers human exposure to environmental factors such as air pollutants and water pollutants. Indicators for air pollution consider outdoor and indoor exposure to $PM_{2.5}$ and use the environmental standards proposed by the World Health Organisation (WHO 2005). Indicators of exposure to water pollution focus on drinking water quality using the standards from relevant legislation (European Council 1998), which is largely based on standards from the World Health Organisation. There are many more substances not covered in the selected indicators that can lead to harmful effects on human health (e.g. persistent organic pollutants, pesticides, etc.), although air pollution and drinking water quality are among the most relevant environmental factors behind health issues. Beyond exposure to chemicals, this category could consider vulnerability to other environmental factors such as extreme weather events, the probability of which is exacerbated by global warming. This type of indicator has not been considered due to its limited responsiveness to policy interventions.

The functions related to amenity and landscape value are represented by standards on the quality of bathing water bodies, the population with nearby green areas next to dwellings and the conservation outlook of relevant World Heritage sites. The former uses concentration of faecal bacteria as the environmental standard (EC 2002). The population with nearby green areas measures access to parks and forests. Last, the indicator of World Heritage sites considers the current state and trend of values, the threats affecting those values, and the effectiveness of protection and management of natural and mixed sites. It should be noted that the indicators selected fall short from covering all non-use values of natural capital, which are not only difficult to capture through indicators, but in many cases also lack science-based environmental standards.

Table 19: Relevance assessment of candidate SES indicators

Function	Principle	Topic	SES indicator	Environmental standard	References
Source	Renew renewable resources	Biomass	Forest utilization rate	Fellings / Net Annual Increment	EEA (2017)
			Fish stocks within safe biological limits	Fishing mortality consistent with Maximum Sustainable Yield Spawning stock biomass consistent with Maximum Sustainable Yield	EC (2010)
		Freshwater	Freshwater bodies not under water stress	Blue water consumption / Mean quarterly flows	Raskin et al. (1997)
			Groundwater bodies in good quantitative status	Good quantitative status as defined in European legislation	EC (2009)
	Use non-renewables prudently	Soil	Area with tolerable soil erosion	Tolerable soil erosion rate	Jones et al. (2004); Huber et al. (2008); Verheijen et al. (2009)
			Area with adequate soil organic matter	Not available	Loveland and Webb (2003)
			Area without land degradation	Non-declining carbon stocks, land productivity and land cover	IAEG-SDGs (2018)
		Abiotic raw materials	Resource-to-production ratio (metal ores)	Not available	-
			Resource-to-production ratio (non-metallic minerals)	Not available	-
			Resource-to-production ratio (fossil fuels)	Not available	-
Sink	Prevent global warming, ozone depletion	Earth system	CO_2 emissions	Long-term CO_2 emissions consistent with a 1.5-2ºC increase in global mean temperature compared to pre-industrial levels.	IPCC (2018)

			ODS consumption	ODS consumption consistent with reducing the ozone hole	UN (1987)
	Respect critical levels and loads for ecosystems	Terrestrial ecosystems	Cropland and forest area exposed to safe ozone levels	Critical levels of tropospheric ozone	Karlsson et al. (2003); Karlsson et al. (2007); Mills et al. (2007)
			Terrestrial ecosystems not exceeding the critical loads of heavy metals	Critical loads of heavy metals	Hettelingh et al. (2015); Hettelingh et al. (2017)
			Terrestrial ecosystems not exceeding the critical loads of eutrophication	Critical load of eutrophication	CLRTAP (2017)
			Terrestrial ecosystems not exceeding the critical loads of acidification	Critical load of acidification	CLRTAP (2017)
		Freshwater ecosystems	Surface water bodies in good chemical status	Good chemical status as defined in European legislation	European Parliament and European Council (2008b)
			Groundwater bodies in good chemical status	Good chemical status as defined in European legislation	EC (2009)
		Marine ecosystems	Marine water bodies in good chemical status	Pollution-related elements of good environmental status as defined in European legislation	EC (2017)
			Plastic pollution	Not available	Villarrubia-Gómez et al. (2018)
Life support	Maintain biodiversity and ecosystem health	Terrestrial ecosystems	Terrestrial habitats in favourable conservation status	Favourable conservation status based on range, area, structure and function.	Röschel et al. (2020)
			Terrestrial area with acceptable biodiversity levels [a]	Local Biodiversity Intactness Index	Steffen et al. (2015b)
		Freshwater ecosystems	Surface water bodies in good ecological status	Good ecological status as defined in European legislation based on	EC (2003)

				biological, physicochemical and hydromorphological parameters	
		Marine ecosystems	Marine water bodies in good ecological status	Good environmental status as defined in European legislation based on biological, physicochemical and hydromorphological parameters	EC (2017)
Human health and welfare	Respect standards for human health	Human health	Population exposed to safe levels of outdoor air pollutants	Critical levels of $PM_{2.5}$, PM_{10} and NO_2	WHO (2005)
			Population exposed to safe levels of indoor air pollutants	Critical levels of PM	WHO (2005)
			Samples that meet safe drinking water criteria	Safe drinking water criteria as defined in European legislation based on microbiological, chemical and other parameters	European Council (1998)
	Conserve landscape and amenity	Other welfare	Recreational water bodies in excellent status	'Excellent' quality criteria as defined in European legislation based on the concentration of Intestinal Enterococci and Escherichia Coli in recreational waters	EC (2002)
			Population with nearby green areas	Green areas that can be reached within 10 minutes' walking.	Poelman (2018)
			Natural and mixed world heritage sites in good conservation outlook	Good conservation outlook based on three elements: the current state and trend of values, the threats affecting those values, and the effectiveness of	Osipova et al. (2014)

				protection and management	

[a]: Biodiversity Intactness Index is often used as a proxy for ecosystem function. This is considered redundant given that ecosystem function is one of the criteria used in the conservation status indicator for terrestrial ecosystems.

Green shading indicates the existence of an environmental standard. Yellow shading indicates the existence of a reference value that is not deemed good enough to be used as an environmental standard. Red shading indicates the lack of environmental standards.

4.2.2.2. Statistical and methodological soundness

Referring to the criteria presented in section 4.2.1.2, Eurostat (2020a, p. 9) recommends "indicators provided by a data producer with a strong commitment to quality, i.e. official statistics or other well established institutions having a quality policy and procedures in place to monitor and report on product quality, will fulfil the above requirements". Here Eurostat's judgment is followed, thereby assuming that the indicators produced by well-established institutions meet the relevant statistical and methodological soundness criteria.

Table 20 shows the agents that have produced the relevant data to populate the indicators. As in the previous case, the indicators are colour-coded to reflect compliance with the soundness criteria. Thus, green shading indicates that the indicator is produced by a well-established institution or institutions associated therewith. Red shading indicates that the indicator currently lacks data.

Most indicators are compiled by European institutions such as the European Environment Agency, the European Commission's Joint Research Centre or by the countries themselves. In this group, we could also consider centres such as the European Topic Centres associated with the European Environment Agency or the European Monitoring and Evaluation Programme of the Convention on Long-range Transboundary Air Pollution. Other indicators are produced regularly by well-established institutions such as Forest Europe and the International Union for Conservation of Nature. The data from the former is used by the European Environment Agency to report on the status of forests.

There are three indicators for which data is not available. In these cases, proxy indicators are proposed to replace them (Table 21). In two cases coastal waters are used to represent marine waters. In the other, an indicator of access to clean cooking fuels is used as proxy for exposure to indoor air pollution. These proxies have adequate environmental standards and therefore meet the relevance criteria.

Table 20: Statistical and methodological soundness assessment of candidate SES indicators

Function	Principle	Topic	SES indicator	Data provider	References
Source	Renew renewable resources	Biomass	Forest utilization rate	Forest Europe	Forest Europe et al. (2015); Forest Europe (2020)
			Fish stocks within safe biological limits	EEA	EEA (2019d)
		Freshwater	Freshwater bodies not under water stress	EEA	EEA (2018c)
			Groundwater bodies in good quantitative status	EEA	EEA (2018b)
	Use non-renewables prudently	Soil	Area with tolerable soil erosion	JRC	Panagos et al. (2015); Panagos et al. (2020)
Sink	Prevent global warming, ozone depletion	Earth system	CO_2 emissions	Eurostat	Eurostat (2019a)
			ODS consumption	UNEP Ozone Secretariat	Ozone Secretariat United Nations Environment Programme (2019)
	Respect critical levels and loads for ecosystems	Terrestrial ecosystems	Cropland and forest area exposed to safe ozone levels	European Topic Centre on Air Pollution, Transport, Noise and Industrial Pollution	Horálek et al. (2019); Horálek et al. (2020)
			Ecosystems not exceeding the critical loads of heavy metals	European Monitoring and Evaluation Programme	Hettelingh et al. (2015)
			Ecosystems not exceeding the critical loads of eutrophication	European Monitoring and Evaluation Programme	Fagerli et al. (2020)
			Ecosystems not exceeding the critical loads of acidification	European Monitoring and Evaluation Programme	Fagerli et al. (2020)
		Freshwater ecosystems	Surface water bodies in good chemical status	EEA	EEA (2018b)
			Groundwater bodies in good chemical status	EEA	EEA (2018b)

		Marine ecosystems	Marine water bodies in good chemical status	Data not available	-
Life support	Maintain biodiversity and ecosystem health	Terrestrial ecosystems	Habitats in favourable conservation status	EEA	EEA (2020a)
		Freshwater ecosystems	Surface water bodies in good ecological status	EEA	EEA (2018b)
		Marine ecosystems	Marine water bodies in good ecological status	Data not available	-
Human health and welfare	Respect standards for human health	Human health	Population exposed to safe levels of outdoor air pollutants	European Topic Centre on Air Pollution, Transport, Noise and Industrial Pollution	Horálek et al. (2019); Horálek et al. (2020)
			Population exposed to safe levels of indoor air pollutants	Data not available	-
			Samples that meet the drinking water criteria	EC	EC (2016)
	Conserve landscape and amenity	Other welfare	Recreational water bodies in excellent status	EEA	EEA (2020c)
			Population with nearby green areas	European Commission	Poelman (2018)
			Natural and mixed world heritage sites in good conservation outlook	International Union for Conservation of Nature	Osipova et al. (2017); Osipova et al. (2020)

Green shading indicates that the data is produced by well-established institutions. Red shading indicates that data is not available.

Table 21: Proxies for SES indicators without data

Function	Principle	Topic	SES indicator	Data provider	References
Sink	Respect critical levels and loads for ecosystems	Marine ecosystems	Marine water bodies in good chemical status	Data not available	-
			Coastal water bodies in good chemical status [a]	EEA	EEA (2018b)
Life support	Maintain biodiversity and ecosystem health	Marine ecosystems	Marine water bodies in good ecological status	Data not available	-
			Coastal water bodies in good ecological status [b]	EEA	EEA (2018b)
Human health and welfare	Conserve landscape and amenity	Other welfare	Population exposed to safe levels of indoor air pollutants	Data not available	-
			Population using clean fuels and technologies for cooking [c]	WHO	WHO (2020)

[a]: The environmental standard is good chemical status as defined in European legislation (European Parliament and European Council 2008b)

[b]: The environmental standard is good ecological status as defined in European legislation based on biological, physicochemical and hydromorphological parameters (EC 2003)

[c]: Members of a household using polluting fuels (e.g. coal, wood, charcoal, dung, crop residues and kerosene) for cooking are considered to be exposed to harmful levels indoor air pollution independent of age and gender that are several times higher than the 24-h exposure guidelines values proposed by WHO (WHO 2018a).

Green shading indicates that the data is produced by well-established institutions. Red shading indicates that data is not available.

4.2.2.3. Data quality

The indicators that passed the soundness test have been assessed against the data quality criteria described in Table 18 (frequency of dissemination, timeliness, geographical coverage, geographical comparability, time coverage and temporal comparability). The results of this assessment are shown in Table 22.

The performance varies considerably across the indicators with those that are reported annually getting high scores in most categories. This is not surprising given that the production of these indicators is well established in the reporting organisations. On the other end, there are indicators that take the form of composite quality indicators. These indicators represent a geographical aggregation of multi-indicator assessments of water bodies or ecosystems and therefore require a lot of data to be produced. For this reason, they are produced every six years. Since the existing compilation guidelines are sometimes implemented differently between countries and since the number of bodies or ecosystems assessed differ between reporting period, their temporal and geographical comparability is sometimes limited.

Out of the 23 indicators that made the previous cut, 22 meet the minimum data quality criteria. Only the indicator on heavy metal critical load exceedance is excluded in this step. The indicator has two data points available, one for the year 2005 and the other is an estimate for 2030 based on modelling results. Thus, the only acceptable data point is too old to be included in the assessment.

As argued previously, the data quality criteria have been relaxed in this thesis because SESI is intended to be a proof of concept. Currently, not all the indicators in Table 22 would meet the stricter criteria that might be necessary if SESI were to be computed regularly and used in policy-making.

Table 22: Data quality assessment of candidate SES indicators

Function	Principle	Topic	SES indicator	Freq.	Time	Gcov	Gcom	Tcov	Tcom
Source	Renew renewable resources	Biomass	Forest utilization rate						
			Fish stocks within safe biological limits						
		Freshwater	Freshwater bodies not under water stress						
			Groundwater bodies in good quantitative status						
	Use non-renewables prudently	Soil	Area with tolerable soil erosion						
Sink	Prevent global warming, ozone depletion	Earth system	CO_2 emissions						
			ODS consumption						
	Respect critical levels and loads for ecosystems	Terrestrial ecosystems	Cropland and forest area exposed to safe ozone levels						
			Ecosystems not exceeding the critical loads of heavy metals						
			Ecosystems not exceeding the critical loads of eutrophication						
			Ecosystems not exceeding the critical loads of acidification						
		Freshwater ecosystems	Surface water bodies in good chemical status						
			Groundwater bodies in good chemical status						
		Marine ecosystems	Coastal water bodies in good chemical status						
Life support	Maintain biodiversity and ecosystem health	Terrestrial ecosystems	Habitats in favourable conservation status						

		Freshwater ecosystems	Surface water bodies in good ecological status						
		Marine ecosystems	Coastal water bodies in good ecological status						
Human health and welfare	Respect standards for human health	Human health	Population exposed to safe levels of outdoor air pollutants						
			Population using clean fuels and technologies for cooking						
			Samples that meet the drinking water criteria						
	Conserve landscape and amenity	Other welfare	Recreational water bodies in excellent status						
			Population with nearby green areas						
			Natural and mixed world heritage sites in good conservation outlook						

Freq: frequency of dissemination; Time: timeliness; Gcov: geographical coverage; Gcom: geographical comparability; Tcov: time coverage; Tcom: temporal comparability.

Green shading represents a score of "high" as defined in Table 18. Yellow and orange shading represent "medium" and "low" scores respectively. Red shading represents an "insufficient" score.

4.2.3. Final indicator set

The final indicator set consists of 21 indicators. While 22 indicators met the minimum data quality criteria, two of the indicators (those on critical loads of eutrophication and acidification in terrestrial ecosystems) have been merged into one through spatial analysis (described below). Each of the 21 indicators shows whether a specific element of natural capital is managed sustainably in that its functioning is not altered in a way that threatens its capacity to provide ecosystem services in the long-term. In order to do so, each indicator is measured against an environmental standard that represents a sustainable reference value. These standards are taken from the scientific literature or from relevant international environmental agreements and EU-level environmental legislation that is informed by expert input. In all cases, the standard has a scientific rationale that links it to good functioning levels. The basic information, including data sources, is also included in Table 23. All indicators and their environmental standards are further described in section 1 of Annex 1.

All in all, there are five indicators for the source function, seven for sink, three for life support and six for human health and welfare. Although at first sight, the difference in the number of indicators assigned to each function might seem striking, it should be noted that some of the indicators in the sink and life support functions are composite metrics of ecosystem condition, each of which consider dozens of parameters. That is the case for those indicators related to ecosystem health and pollution (e.g. conservation status of terrestrial ecosystems, and the chemical and ecological status of water bodies). The exception would be the chemical status of terrestrial ecosystems. Since it was not possible to generate a single composite metric for this one, two separate indicators have been used: one for ozone pollution and one for eutrophication and acidification. The latter is the result of spatially aggregating with the one-out-all-out rule the critical load exceedance maps for eutrophication and acidification.

Because of the different geographical contexts and natural resource endowments, all the indicators do not have the same importance for all the countries. While this issue could be partly dealt with through weighting, there are more extreme cases that require some indicators to be excluded when computing the index for some countries. This is, for instance, the case of marine waters in countries that do not have access to the coast.

Table 23: Final SES indicator set

Function	Principle	Topic	SES indicator [Unit]	Data	Standard	References
Source	Renew renewable resources	Biomass	Forest utilization rate [%]	Forest Europe et al. (2015); Forest Europe (2020)	Fellings / Net Annual Increment	EEA (2017)
			Fish stocks within safe biological limits [%]	EEA (2018a, 2019d)	Fishing mortality consistent with Maximum Sustainable Yield Spawning stock biomass consistent with Maximum Sustainable Yield	EC (2010)
		Freshwater	Freshwater bodies not under water stress [%]	EEA (2018c)	Blue water consumption / Mean quarterly flows	Raskin et al. (1997)
			Groundwater bodies in good quantitative status [%]	EEA (2018b)	Good quantitative status as defined in European legislation	EC (2009)
	Use non-renewables prudently	Soil	Area with tolerable soil erosion [%]	Panagos et al. (2015); Panagos et al. (2020)	Tolerable soil erosion rate	Jones et al. (2004); Huber et al. (2008); Verheijen et al. (2009)
Sink	Prevent global warming, ozone depletion	Earth system	CO_2 emissions [tonnes per capita]	Eurostat (2019a)	Long-term CO_2 emissions consistent with a 1.5-2ºC increase in global mean temperature compared to pre-industrial levels.	IPCC (2018)
			ODS consumption [tonnes per capita]	Ozone Secretariat United Nations Environment Programme (2019)	ODS consumption consistent with reducing the ozone hole	UN (1987)
	Respect critical levels and loads for ecosystems	Terrestrial ecosystems	Cropland and forest area exposed to safe ozone levels [%]	Horálek et al. (2019); Horálek et al. (2020)	Critical levels of tropospheric ozone	Karlsson et al. (2003); Karlsson et al. (2007); Mills et al. (2007)
			Terrestrial ecosystems not exceeding the critical loads of eutrophication and acidification [%]	Fagerli et al. (2020)	Critical loads of eutrophication and acidification	CLRTAP (2017)
		Freshwater ecosystems	Surface water bodies in good chemical status [%]	EEA (2018b)	Good chemical status as defined in European legislation	European Parliament and

						European Council (2008b)
			Groundwater bodies in good chemical status [%]	EEA (2018b)	Good chemical status as defined in European legislation	EC (2009)
		Marine ecosystems	Coastal water bodies in good chemical status [%]	EEA (2018b)	Pollution-related elements of good environmental status as defined in European legislation	EC (2017)
Life support	Maintain biodiversity and ecosystem health	Terrestrial ecosystems	Terrestrial habitats in favourable conservation status [%]	EEA (2020a)	Favourable conservation status based on range, area, structure and function.	Röschel et al. (2020)
		Freshwater ecosystems	Surface water bodies in good ecological status [%]	EEA (2018b)	Good ecological status as defined in European legislation based on biological, physicochemical and hydromorphological parameters	EC (2003)
		Marine ecosystems	Coastal water bodies in good ecological status [%]	EEA (2018b)	Good environmental status as defined in European legislation based on biological, physicochemical and hydromorphological parameters	EC (2017)
Human health and welfare	Respect standards for human health	Human health	Population exposed to safe levels of outdoor air pollutants [%]	Horálek et al. (2019); Horálek et al. (2020)	Critical levels of $PM_{2.5}$	WHO (2005)
			Population using clean fuels and technologies for cooking [%]	WHO (2020)	Critical levels of $PM_{2.5}$	WHO (2005)
			Samples that meet the drinking water criteria [%]	EC (2016)	Safe drinking water criteria as defined in European legislation based on microbiological, chemical and other parameters	European Council (1998)
	Conserve landscape and amenity	Other welfare	Recreational water bodies in excellent status [%]	EEA (2019f)	'Excellent' quality criteria as defined in European legislation based on the concentration of Intestinal Enterococci and Escherichia Coli in recreational waters	EC (2002)
			Population with nearby green areas [%]	Poelman (2018)	Green areas that can be reached within 10 minutes' walking.	Poelman (2018)

			Natural and mixed world heritage sites in good conservation outlook [%]	Osipova et al. (2017); Osipova et al. (2020)	Good conservation outlook based on three elements: the current state and trend of values, the threats affecting those values, and the effectiveness of protection and management	Osipova et al. (2014)

4.2.4. Structure of SESI

The 21 SES indicators have been arranged in several layers as shown in Figure 6. These layers represent the functions, sustainability principles and topics included in previous tables and serve as the dimensions across which the results will be aggregated to generate the final index score.

Figure 6: Structure of SESI

Note for small figure: I (index), F (function), P (principle), T (topic), ind (indicator).

4.3. Data treatment

4.3.1. Data gaps

Table 24 shows the availability of data for each of the indicators that comprise SESI. The table shows the cases for which data is not available and for which the indicator does not apply. The latter refers to landlocked countries without access to coastal and marine waters, and countries that do not have natural or mixed world heritage sites within their borders. The table also shows the latest year for which data is available. This is the reference year used to compute SESI. As with other indices, the reference year used in the underlying indicator differs. In a few cases, the reference year lies in the 2010-2013 period, which is not ideal, but is considered here sufficient to showcase the potential of SESI as a single measure of environmental sustainability.

Table 24: Data gaps in SES indicators

SES indicator	Year	Available	Gaps	Not applicable	% available ([a])
Forest utilization rate	2015	20	7	1	74
Fish stocks within safe biological limits	2017	28	0	0	100
Freshwater bodies not under water stress	2015	27	1	0	96
Groundwater bodies in good quantitative status	2015	28	0	0	100
Area with tolerable soil erosion	2016	28	0	0	100
CO_2 emissions	2018	28	0	0	100
ODS consumption	2019	28	0	0	100
Cropland and forest area exposed to safe ozone levels	2017	28	0	0	100
Ecosystems not exceeding the critical loads of eutrophication and acidification	2018	28	0	0	100
Surface water bodies in good chemical status	2015	28	0	0	100
Groundwater bodies in good chemical status	2015	28	0	0	100
Coastal water bodies in good chemical status	2015	23	0	5	100
Habitats in favourable conservation status	2018	28	0	0	100
Surface water bodies in good ecological status	2015	27	1	0	96
Coastal water bodies in good ecological status	2015	23	0	5	100
Population exposed to safe levels of outdoor air pollutants	2017	28	0	0	100
Population using clean fuels and technologies for cooking	2018	27	1	0	96
Samples that meet the drinking water criteria	2013	27	1	0	96
Recreational water bodies in excellent status	2019	28	0	0	100
Population with nearby green areas	2012	28	0	0	100
Natural and mixed world heritage sites in good conservation outlook	2020	20	0	8	100

([a]): Availability is computed excluding the countries for which the indicator is not applicable.

When a country does not report data in one of the indicators above (see the column 'gaps'), data can be imputed to fill the gaps. Ignoring the missing values is the simplest method to treat the data. Nonetheless, ignoring missing values is equivalent to undertaking a shadow imputation, which in practice means assuming that the missing value is equal to the mean of the indicators in the (sub)dimension. For instance, in a subdimension that comprises three indicators for which two values are available (e.g. population exposed to acceptable outdoor pollution = 50 and population using clean cooking fuels and technologies = 100) and one value is missing (e.g. samples that meet the drinking water criteria), when aggregating at higher levels with an arithmetic mean, the value of the subdimension (e.g. human health) would be 75 ((50+100)/2). This is equivalent to assigning a value of 75 to the indicator for which no data was available ((50+100+75)/3). Thus, ignoring the missing values assigns a shadow value and therefore affects the final score of the index. Considering this, the following stepwise data treatment strategy has been implemented to (whenever possible) estimate more realistic values than those that would be obtained through shadow imputation:

- If the country has data for a previous year and the indicator is (at least moderately) correlated (R>0.4) with a related indicator in that same year, the same temporal change as the correlated indicator is assumed in order to produce a value for the year indicated in Table 24.
- In the absence of strongly correlated indicators, if data is available for a previous year, that data point is used. This is not an imputation, but the use of a different year as reference.
- If the data for the indicator is (at least moderately) correlated (R>0.4) with that of a different indicator, linear regression is used to estimate the missing value.
- As a last option, the average of the remaining countries is assigned to the country.

This approach results in imputations shown in Table 25.

Table 25: Approach to impute data in SES indicators

Indicator	Approach	Description
Forest utilization rate	Same temporal change as correlated indicator.	See section 1 of Annex 1.
Freshwater bodies not under water stress	Linear regression with correlated indicator.	Data for Cyprus estimated based on the correlation between this and annual Water Exploitation Index +.
Surface water bodies in good ecological status	Linear regression with correlated indicator.	Data for Malta estimated based on the correlation between this and "Coastal water bodies in good ecological status".
Population using clean fuels and technologies for cooking	Data from previous year.	2016 data used for 2018 in Bulgaria.
Samples that meet the drinking water criteria	Country average.	Croatia was not part of the EU in 2013, the latest year for which data is available. The average from the other countries is taken.

4.3.2. Outliers

Treating outliers should be discussed on a case-by-case basis. Arguably, outliers should be treated when they represent a measure or encoding mistake, since they can be problematic when the normalisation process depends on the sample distribution (Becker et al. 2019). No such cases have been identified in the SES indicators. Since the normalised scores depend on meeting predefined environmental standards and not on the sample distribution, treating outliers offers no benefits and would represent a departure from measured reality.

4.4. Normalisation

Most of the indicators in an index usually have different units, which makes them incomparable unless transformed into a common unitless scale. This is the goal of the normalisation process. There are multiple normalisation methods (OECD and JRC 2008), so the selection of a method is not trivial. The relevance of environmental standards in the conceptual framework of SESI demands the goalpost method to be used in the normalisation process. In this method, user-defined values are used as goalposts (i.e. upper and lower bounds) to transform indicators into a scale between 0 and 100. For the normalisation process to be aligned with the strong sustainability narrative, these upper and lower bounds need to represent full and no compliance of environmental standards respectively.

The normalised scores are calculated as shown in equation below, where the normalised value of an indicator (NI) depends on the value of the indicator (I), and values assigned as goalposts (gp_{min} and gp_{max}). Thus, gp_{min} represents no compliance with environmental standards and therefore leads to a normalised score of 0. Conversely, gp_{max} represents full compliance with the environmental standard, thereby leading to a normalised score of 100. Normalised scores lower than 0 and higher than 100 are assigned 0 and 100 values.

Equation 1

$$NI = 100 \frac{I - gp_{min}}{gp_{max} - gp_{min}}$$

Most of the 21 indicators selected describe environmental or social states as percentages of ecosystems, bodies or population that meet environmental standards. Thus, the values all fall between 0 and 100 and are therefore implicitly normalised, where in all the cases 0 is the worst possible performance and 100 the best.

A few indicators (forest utilization rate, per-capita GHG emissions and per-capita ODS consumption) are interpreted differently, since the values are not bound in the 0-100 range as the previous indicators. GHG emissions and ODS consumption are pressure instead of state indicators as defined in the DPSIR framework (EEA 2003b) and in theory their values can go from -∞ to +∞ when considering negative emissions and the destruction of ODS. Forest utilisation rate, on the other hand, is a state indicator, but does not describe a percentage of forests that comply with an environmental standard. Its values can range from 0 to +∞. In these cases, the goalpost values (gp_{min} and gp_{max}) of each indicator are shown in Table 26 and justified in the corresponding indicator fiche in section 1 of Annex 1. As explained before, gp_{min} and gp_{max} represent the values that would lead to normalised scores of 0 and 100 respectively. In the cases below, gp_{min} values are higher than gp_{max} values. A worked example of the normalisation process is given in Table 37 in Annex 1.

Table 26: Normalisation equations used for SES indicators

Function	Principle	Topic	SES indicator [Unit]	gp_{min}	gp_{max}
Source	Renew renewable resources	Biomass	Forest utilization rate [%]	100	70
Sink	Prevent global warming, ozone depletion	Earth system	CO_2 emissions [tonnes per capita] ([a])	2.5	0.5
			ODS consumption [tonnes per capita] ([b])	0.00032	0

([a]): gp_{max} shows the per-capita CO_2 emissions consistent with meeting the 1.5ºC target with 67% of possibilities based on the carbon budgets of IPCC (2018). On the other hand, gp_{min} is consistent with meeting the 2ºC target with 33% of possibilities. Emissions have been allocated on an equal-per-capita basis using cumulative population figures.

([b]): gp_{min} shows the per capita consumption of ODS in 1989, which represents the peak of the destruction of the ozone layer.

4.5. Weighting

The weights assigned to the indicators that will be aggregated is a reflection of their importance, yet this does not necessarily represent how much they impact the final score (Becker et al. 2017).

Reconciling the theoretical framework with weighting is a particularly problematic process in the case of SESI. The issue of weighting can be approached from two perspectives. The first one refers to the weights assigned to each of the dimensions of the index; in this case functions, sustainability principles, topics and indicators. The second one relates to the use of the same set of weights across countries.

When it comes to the weights of the dimensions, life support functions should take preference over source, sink and human health and welfare functions because without life support functions, the other functions would not be able to be sustained in the long-term. At lower levels, prioritising sustainability principles becomes more difficult. For instance, in the source functions, the relevance of renewable and non-renewable resources depends on the domestic endowments. In sink functions, prioritising global vs regional pollution neutralisation processes is not straightforward. In the case of human health and welfare functions, human health should come before the functions related to other aspects of welfare. At the level of indicators, it becomes almost impossible to assign weights based on relevance, since different natural capital endowments and the uneven contribution of pollutants to overall environmental and health impacts differ considerably.

Regarding country weights, maximising the national policy impact of the index would warrant country-specific weights for the elements of natural capital that are adapted to the context of each country. Nonetheless, this would render the results between countries incomparable, thereby reducing the potential use of SESI in a global context.

Translating the arguments provided in the paragraphs above into weights remains problematic. After all, it would require expert input or a criterion that could generate a broad consensus, e.g. some kind of valuation exercise that reflects the real value (monetary or otherwise) of the ecosystem services provided by natural capital to people, or mortality rates attributable to different pollutants, or even historical responsibility for activities generating long-lasting impacts (e.g. carbon emissions). In the absence of any weighting method likely to generate broad agreement, equal weights are here assigned to all the indicators from top to bottom. Thus, each of the four functions of natural capital has a weight of 0.25, while the principles therein have weights of 0.5, except in the case of the maintenance of biodiversity and ecosystem health, which has a weight of 1 because it is the only principle assigned to the life support functions. This logic is applied to the topics and the SES indicators. It should be noted that as pointed out by Hsu et al. (2013), the weighting process is as much a political process as it is a scientific process. As a result, it can be easily challenged irrespective of the method used. The use of equal weights ensures comparability across countries' scores and still leaves room for alternative weighting approaches in separate exercises where the results are expected to be used in country-specific settings.

4.6. Aggregation

As in previous steps (although in the case of weighting no alternative scheme to equal weighting seemed justifiable), the aggregation process should be formulated along the lines of the theoretical framework. The concepts of 'strong sustainability' and 'critical natural capital' are at the core of the ESGAP framework. In combination, both concepts address the substitution capacity between natural capital and other types of capital, as well as between the different functions of natural capital. The limited substitutability between the different types of capital is reflected in that SESI is an independent metric of environmental sustainability, addressing the environmental dimension of sustainable development irrespective of broader social and economic issues. The limited substitutability between the functions of natural capital, on the other hand, is reflected in the type of mean used in the aggregation process.

Slightly adapting the formulation of Rickels et al. (2016), the final score of a composite indicator *CI* is calculated through the generalised mean, which is expressed as follows.

Equation 2

$$CI = \left(\sum_{i=1}^{n} \alpha_i NI_i^{\rho}\right)^{\frac{1}{\rho}}$$

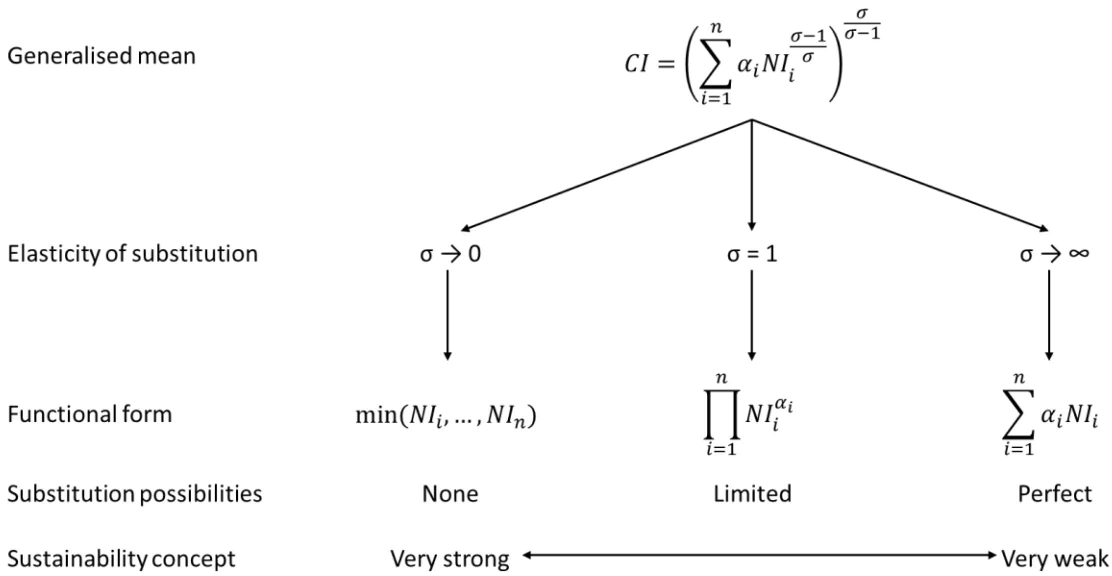
In the previous equation, *n* represents the sum of *i* indicators, α represents the weights and ρ the substitution possibilities between indicators. In this vein, the parameter ρ is a function of the elasticity of substitution σ.

Equation 3

$$\rho = \frac{\sigma - 1}{\sigma} \quad \text{with} \quad 0 \leq \sigma \leq \infty$$

The value of σ determines the extent to which the functions addressed by the indicators can be substituted for each other and therefore defines where the index lies in the very weak to very strong sustainability continuum. In this context, when functions are assumed to be completely interchangeable (with high values of σ), the index is calculated as the weighted arithmetic mean of the normalised indicators. With the weighted arithmetic mean, poor performance in one dimension is linearly compensated for by high achievement in another dimension and therefore it implicitly assumes that the functions provided by natural capital are interchangeable. In the opposite case, in absence of any substitution possibilities (when σ is close to 0) the index takes the value of the normalised indicator with the lowest score. Thus, the performance of a country would equal its performance in the worst dimension when weights are equal. This rule could be used to aggregate indicators spatially, rather than through individual scores, similar to the one-out all-out rule, but it does not seem useful to compute an index. In between we can find assumptions of imperfect substitution. For instance, when σ takes a value of 1, the mean takes the form of a geometric mean. With a weighted geometric mean, low scores in any dimension are directly reflected in the final composite indicator. This is shown in Figure 7.

Figure 7: Generalised mean in the context of weak and strong sustainability



Source: Adapted from Rickels et al. (2016)

The strong sustainability proposition is built around the notion of limited substitution. In a nested structure such as the one of SESI, different substitution elasticities can be applied at different layers and within layers. Nonetheless, ground-truthing these values is not possible, which makes the choices of elasticities a normative process. Considering the audience of SESI and the arbitrariness of selecting substitution elasticities, the geometric mean seems to be the most reasonable option to reflect limited substitution capacity, although it is harder to understand than a simple arithmetic mean.

Nonetheless, the use of the geometric mean in some contexts also has its drawbacks. In this case, the main limitation of the geometric mean is that it collapses to zero when any indicator has a value of zero. For SESI, the normalised scores were re-scaled by replacing the values below a lower bound by the value of the lower bound to avoid the presence of zeros, similar to other indices (see Box 4). Integers from one to five were tested to be used as a lower bound of the normalised scores. While the choice of these values is arbitrary, a consistency check was undertaken to validate the choice. Thus, country rankings obtained using the Copeland rule[10] were compared to the rankings based on SESI scores calculated after replacing small normalised values by the integers in the 1-5 range. The value five performed the best and therefore was adopted as a lower constraint to correct zeros and small values. Full details are given in section 3 in Annex 1.

Box 4: Treatment of zeros when using the geometric mean to aggregate in other indices

Several researchers prefer to use a geometric mean in the aggregation process to represent the limited substitutability potential between indicators and between dimensions of indices. Since the geometric mean requires positive values for the index score to avoid collapsing to zero, the zeros in the sample need to be treated.

[10] The Copeland rule is a pairwise-voting method that ranks countries based on a scoring system of pairwise 'wins', 'losses', and 'ties'. In the case of SESI, each country's performance in the 21 indicators has been compared to that of the other 27 countries. The countries have been ranked based on the number of wins, ties and losses of each country.

When the zeros are the result of values below the detection limit of the device used for the measurement in a lab, there are different strategies to treat them (Helsel 2005). Nonetheless, this is not the case in SESI, since absolute zeros are possible when using indicators that show whether a percentage of ecosystems or population meets a given environmental standard. In this context, several workarounds have been proposed to overcome the presence of zeros in the calculation of the geometric mean, none of which is exempt of criticism. Common solutions are adding small values to the zeros (Martín-Fernández et al. 2003; O'Brien et al. 2010) or using the Williams mean (Williams 1937), which adds a value of one to every measurement and subtracts it from the geometric average (see references in de la Cruz and Kreft (2018).

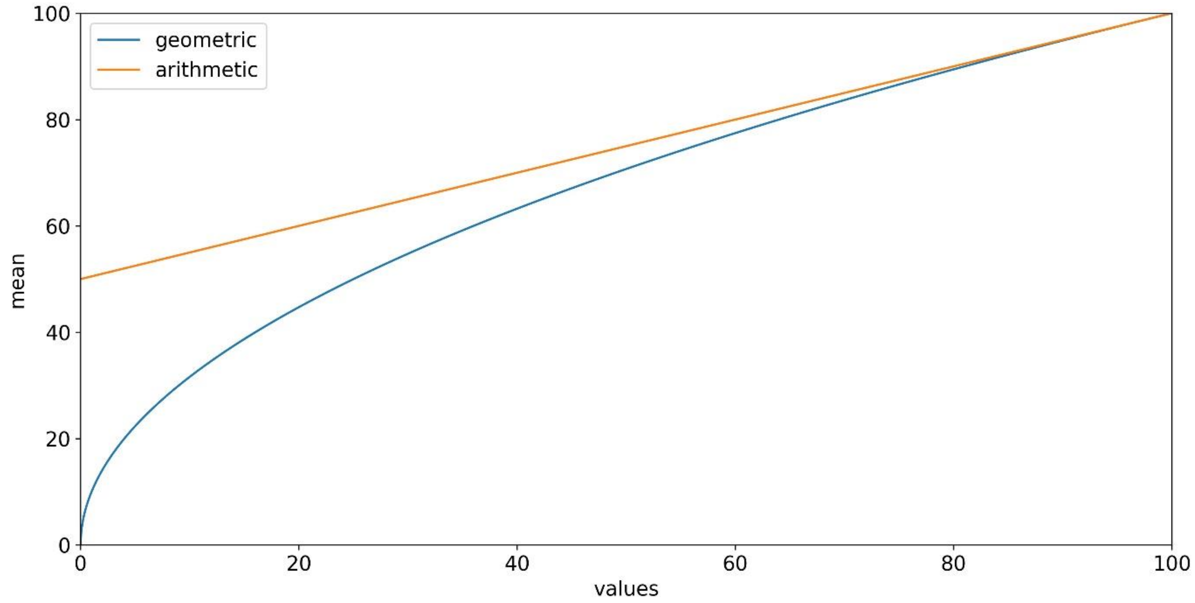
Table 27 summarises the approach taken in selected indices. Although slightly different approaches are taken, all the examples set minimum values to avoid the presence of zeros.

Table 27: Treatment of zeros in different indices

Index	**Approach**	**Reference**
Inequality-adjusted Human Development Index	For mean years of schooling one year is added to all valid observations to compute the inequality. For income per capita, negative and zero incomes and incomes in the bottom 0.5 percentile are replaced with the minimum value of the second bottom 0.5 percentile of the distribution of positive incomes.	UNDP (2020)
Gender Inequality Index	A minimum value of 0.1 percent is set for all indicators.	UNDP (2020)
Sustainable Society Index	A minimum value of 1 percent is set for all indicators.	Sironen et al. (2015)
Europe 2020 Index	A minimum value of 0.1 percent is set for all indicators.	Pasimeni (2013)
Rule of Law Index	The data was re-scaled onto a 1-100 range when using the geometric mean in sensitivity analysis.	Saisana and Saltelli (2011)

The value chosen is arbitrary to some extent and has an implicit impact on the assumption of substitutability between the functions of natural capital when one of the functions is threatened (i.e. when the normalised score of an indicator is close to zero). Figure 8 shows the geometric mean of two indicators where one has a score of 100 and the other one the score represented in the X axis. The differences are not trivial. For instance, the geometric mean of 100 and 0.01 equals one, while the geometric mean of 100 and one equals 10. Thus, adopting 0.01 as lower bound to re-scale the normalised values assumes that there is less substitution capacity between the functions of natural capital. This issue is further explored through uncertainty analysis in section 4.8 and the implications further discussed in section 4.10.

Figure 8: Comparison between geometric and arithmetic means



The figure shows the geometric and arithmetic mean values between a perfect score of 100 and the values in the x axis.

4.7. Statistical and conceptual coherence

Statistical and conceptual coherence are relevant attributes of an index. A statistical coherence analysis can show the extent to which the results of an index capture the information included in the underlying indicators (e.g. by highlighting redundant indicators or information loss in the aggregation process). Statistical coherence can be analysed through correlation analysis, principal component analysis and similar tools (JRC 2019). On the other hand, conceptual coherence is a more qualitative attribute of an index that reflects its consistency with the theoretical framework. Both analyses are commonly used to revise choices made during the selection of indicators, their allocation to (sub)dimensions of the index, or choices related to the normalisation, weighting or aggregation processes.

4.7.1. Conceptual coherence

Consistency with the theoretical framework is key in SESI, especially considering that the index is presented as a proof of concept intended to promote strong sustainability thinking at the national level. There are several decisions along the process of building the index that are intended to reflect specific features of the theoretical framework. These are summarised in Table 28.

The indicator selection process is aligned with the concepts of strong sustainability and environmental sustainability through the ‘relevance’ criterion. In order to reflect that the limited substitutability between the functions of natural capital and other types of capital, SES indicators – and by extension SESI – are only related to the former. Likewise, in line with the environmental sustainability definition, SES indicators require a science-based environmental standard against which performance can be measured. The use of environmental standards is also key part of the normalisation process to set upper and lower bounds with the goalpost method. In the case of environmental and social state indicators, the environmental standards are part of the indicator used in that the indicator measures the percentage of population, ecosystems or similar variable that meets the environmental standard. The upper and lower bounds are set as no and full compliance respectively. In a few cases, the goalpost method requires the environmental standards to be set as upper and lower bounds. Other normalisation methods such as the min-max and z-scores assign normalised values based on the relative performance of countries and therefore depend on the sample distribution.

The theoretical framework argues that life support functions are more relevant than the source, sink and human health and welfare functions. While that could be represented by assigning different weights to the functions, the lack of a suitable method prevents doing so. As a result, provisionally equal weights are assigned to all the dimensions and indicators of the index. This aspect is considered in the uncertainty analysis.

On the issue of substitutability between the functions of natural capital, choosing between the arithmetic and geometric means, and the Leontief production function determines whether full, limited or no substitution capacity respectively is assumed between the functions represented by the indicators. In this case, the geometric mean is the one that is more closely aligned with the concept of strong environmental sustainability. Nevertheless, while the geometric mean penalises low performances, it collapses to zero with very small values, which would indicate that at such levels the substitution capacity is null. Replacing those values is a common method used to avoid that problem.

Table 28: Conceptual coherence assessment of SESI

Framework	Process	Description
Strong sustainability assumes limited substitution capacity between the functions of natural capital and other types of capital	Indicator selection	The first criterion in the indicator selection is relevance. This requires the indicators to be related to the functions of natural capital.
Environmental sustainability should be represented through environmental standards.	Indicator selection	The first criterion in the indicator selection is relevance, which requires the indicators to have science-based environmental standards.
	Normalisation	The goalpost method requires upper and lower bounds to be defined to normalise the indicators. These are set as full or no compliance with environmental standards.
Some environmental functions are more relevant than others	Weighting	Due to the lack of a suitable method, equal weights are assigned to all the dimensions and indicators.
Strong sustainability assumes limited substitution capacity between the functions of natural capital	Aggregation	The geometric mean penalises low scores, which can be interpreted as limited substation capacity.
	Treatment of zeros	The presence of zeros makes the geometric mean collapse to zero, therefore implying that there is no substitution capacity. Treating zero and low values avoids this problem.

4.7.2. Statistical coherence

Statistical coherence is commonly used to assess the transfer of information from the indicators to the index through the different dimensions and subdimensions. In order to maximise the information transfer, the indicators should be positively correlated with the subdimensions to which they have been allocated, the latter should be positively correlated with their corresponding dimension, etc. This should happen without having collinear indicators, since this leads to redundancy and overweighting of certain phenomena

(Papadimitriou et al. 2020). The statistical coherence of SESI has been assessed through two cross-correlation analyses. One between the indicators and their corresponding dimensions in the index, and the other one between the different dimensions in the index. Overall, the results, which are described in detail in section 4 of Annex 1, suggest that the index should not be interpreted on its own and therefore should be complemented with the function scores to ensure that limited information is lost in the aggregation process.

4.8. Uncertainty analysis

The construction of an index requires making assumptions related to indicator selection, data treatment, normalisation, weighting and aggregation and thus, understanding the effects of the choices made is critical to properly interpreting the results. While one could test the effects of every single assumption made, a more targeted approach is proposed here, which focuses on how the theoretical framework is reflected in the selection of the normalisation, weighting and aggregation methods. The assumptions tested are described in Table 29 and explained in the following paragraphs. The table shows the elements of the theoretical framework the assumption is related to, the method used to represent that element and the alternative assumptions tested.

Table 29: Assumptions tested in the uncertainty analysis of SESI

Process	Default	Test
Normalisation	Goalpost	Min-max
Weighting	Equal weights	Life support is more relevant Life support is critical
Aggregation	Geometric mean	Arithmetic mean Minimum score

One of the key elements of the ESGAP framework is the notion that environmental sustainability requires measuring absolute, not relative performances. Absolute performance can only be measured against science-based environmental standards, as opposed to measuring country performance relative to their peers (relative performance). Environmental sustainability requires the normalisation of indicators through the goalpost method where compliance with environmental standards is reflected in the goalpost values chosen. Other indices use the min-max method to normalise, where scores reflect the position of a given country compared to the best and worst performers in the sample, which is indicative of relative instead of absolute performance. In order to assess the effects of the normalisation method, the results obtained at index and function level were compared to those obtained by calculating SESI using the min-max normalisation method as shown in Equation 4. I_{min} and I_{max} values are calculated as the 2.5th and 97.5th percentiles of the values of each indicator across the 28-country sample.

Equation 4

$$NI = 100\frac{I - I_{min}}{I_{max} - I_{min}}$$

The weights assigned to the dimensions and indicators that form an index are intended to reflect their relevance. Because of the lack of an agreed method to capture the differences in how the natural capital functions contribute to human welfare, equal weights have been assigned to all the dimensions and indicators. Nonetheless, the theoretical framework argues that life support functions underpin life on Earth and therefore the other functions cannot exist without the former. Two uncertainty tests are undertaken in this regard. In the first one, the weights of the source, sink, life support and human health and welfare functions are changed to reflect the position that life support functions are more relevant than the rest. In this case, the weights have been set to 0.4 for life support functions, and 0.2 for the source, sink, and human health and welfare functions. The second hypothesis is that not only are life support functions more relevant, but critical. To reflect this, life support functions have been assigned a weight of 0.7, compared to 0.1 in the other functions.

The choice of the aggregation method reflects the stand taken in the strong-weak sustainability continuum. This is related to the substitution capacity between the functions of natural capital and other types of capital (e.g. manufactured, human, social), and within the functions of natural capital itself (source, sink, life support, and human health and welfare). As argued in section 4.6, the arithmetic and geometric means represent the weak and strong sustainability positions, while the value of the lowest-scoring indicators would represent the very strong sustainability position defined by Turner (1993) (i.e. lack of substitution capacity). SESI uses the geometric mean to aggregate the information from the lower levels of the index, thereby taking a strong sustainability position. Sensitivity to the arithmetic mean and the Leontief production function (i.e. the minimum indicator value) are tested. In this context, the minimum value adopted in the normalisation process to avoid non-zero values when using the geometric mean also reflects the substitution capacity between the functions of natural capital when at least one of those functions is severely impaired. For instance, the geometric mean of two equally weighted indicators with the scores 5 and 100 is 22.4, while the geometric mean of the scores 1 and 100 is 10. Sensitivity to this assumption is also tested.

The different normalisation, weighting and aggregation options described above are tested separately using Pearson and Spearmann correlations. These methods show the linear relationship of the results and their ranking respectively. The purpose of analysing the uncertainty to the normalisation, weighting and aggregation methods separately is to shed light on how assumptions made to reflect specific elements of the ESGAP framework affect the results of SESI. The insights might be useful to understand how choices made in other environmental indices that might not be fully aligned with their theoretical framework could significantly affect the main messages derived from those indices.

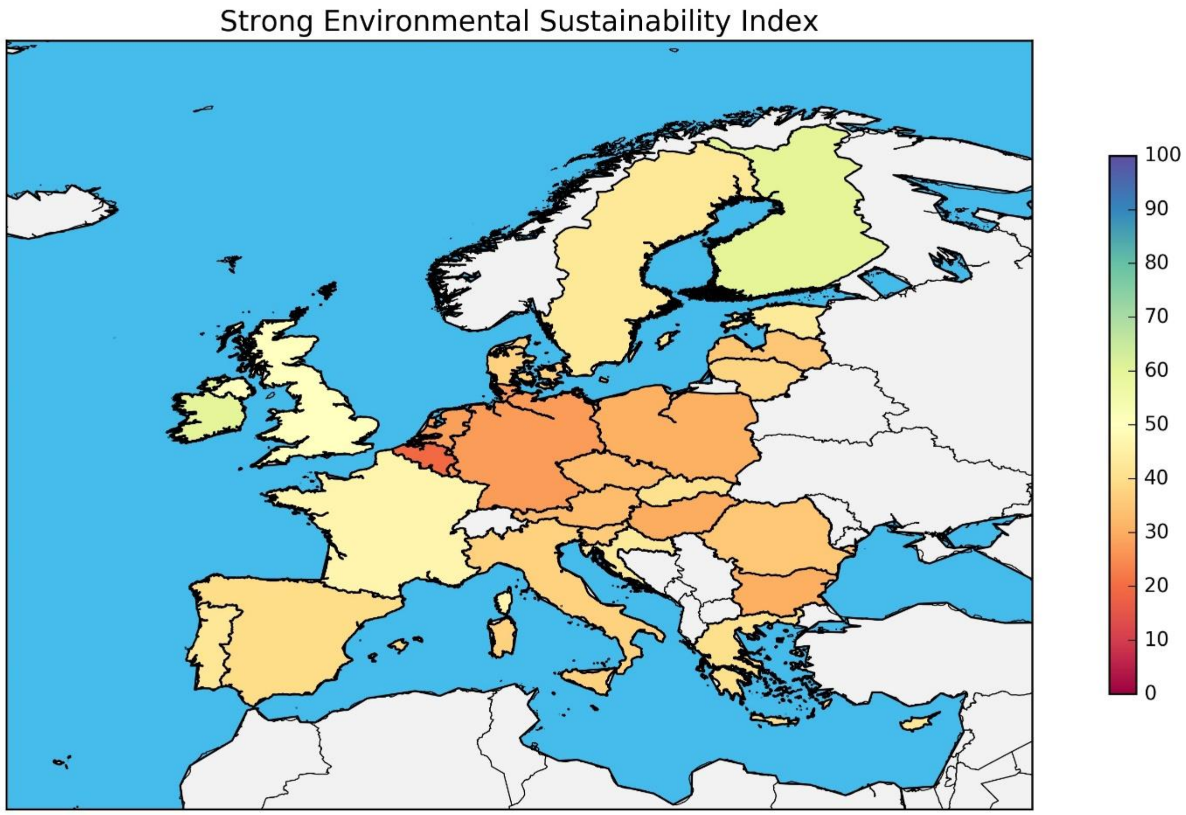
4.9. Results

Figure 9 shows the index score of European countries according to their most recent data point. The same information is shown in Figure 10, with the countries sorted based on their score. For consistency, this order is maintained in the following figures when displaying the results.

The Anglo-Celtic isles and the Scandinavian countries seem to perform better than the Mediterranean, and central and eastern European countries. Nonetheless, the absolute scores are low in most cases, suggesting that one or more environmental functions are currently jeopardised in many countries. Only three countries score more than 50 points and the maximum score is 60, which is obtained by Finland. After the frontrunners, 18 countries obtain scores between 30 and 45, while six countries score lower than 30, with

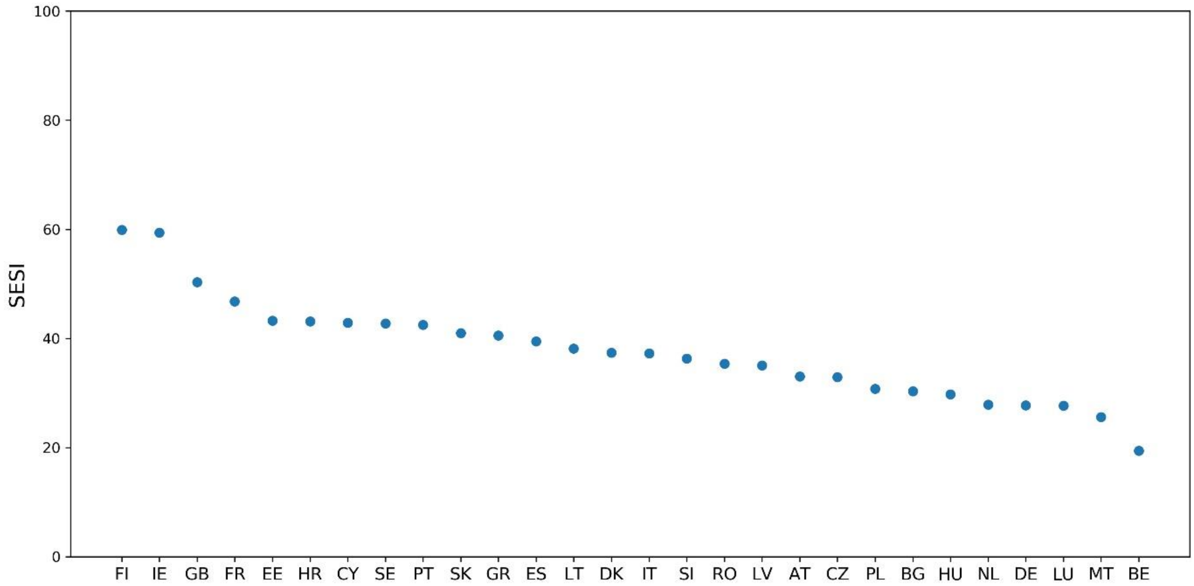
Belgium being at the bottom with 19 points. When considered as a block, Europe gets a score of 47. Of course, at the index level, the score is influenced by the use of the geometric mean in the aggregation, since this penalises low performances in individual indicators. Thus, countries that perform poorly in several indicators will see their aggregate score reduced, thereby reflecting the limited substitution capacity between the environmental functions represented by the indicators.

Figure 9: SESI score for European countries



SESI scores countries from 0 to 100 in terms of their environmental sustainability performance. A score of 100 indicates the compliance of all the indicators across the four environmental functions with their corresponding environmental standard. A score of 0 indicates the opposite.

Figure 10: SESI score for European countries



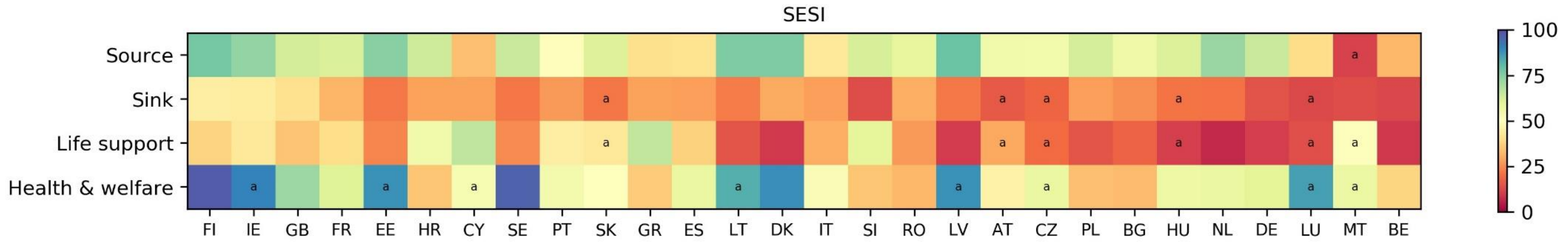
SESI scores countries from 0 to 100 in terms of their environmental sustainability performance. A score of 100 indicates the compliance of all the indicators across the four environmental functions with their corresponding environmental standard. A score of 0 indicates the opposite. Countries are sorted by the total index score from higher to lower.

As with any index, the total score can hide disparities in the performance at lower levels of aggregation. In this context, Figure 11 and Figure 12 show country scores for the four broad environmental functions and the seven sustainability principles used to characterise environmental sustainability. Countries perform very differently in source, and human health and welfare functions, with countries in the first positions scoring relatively high in those two functions. In the source function, which cover the provision of forest and fish biomass, surface and groundwater, and soil, former Soviet Union and Scandinavian countries hold the first five positions with scores over 70. Most countries obtain scores between 40 and 65. Europe as a whole sits at the upper side of the range with a score of 62. Former Soviet Union and Scandinavian countries, as well as the Anglo-Celtic isles are the frontrunners in the human health and welfare function. Countries such as Finland, Sweden and Ireland score over 90. This means that these countries almost comply with the science-based standards used for (indoor and outdoor) air pollution, drinking water, bathing waters, access to green spaces and the conservation of relevant World Heritage sites. The European block obtains a score of 64 in this category.

The sink and life support functions describe a different picture. Scores are more homogeneous with almost every country performing poorly. In the case of sink functions, none of the countries reaches 50 (Europe scores 33). 23 countries have scores below 30. In the countries with the highest scores, the main explanatory factor is the poor performance in CO_2 emissions, where none of the countries are in the sustainable emission range (0.5-2.5 t CO_2 per capita). In the remaining countries, poor performance in CO_2 emissions is combined with poor performance in terrestrial, freshwater or coastal ecosystems' pollution. Scores in life support functions are also generally low with 23 countries getting scores under 50. Five Mediterranean countries (Cyprus, Greece, Slovenia, Croatia and Malta) are the top performers with scores up to 66. This is driven by high scores in the ecological condition of their coastal ecosystems. It should be noted that this indicator is used as a proxy for marine ecosystems, which currently lacks data.

Given that Mediterranean fish stocks are largely overexploited, it seems unlikely that countries will report such good conditions in marine ecosystems.

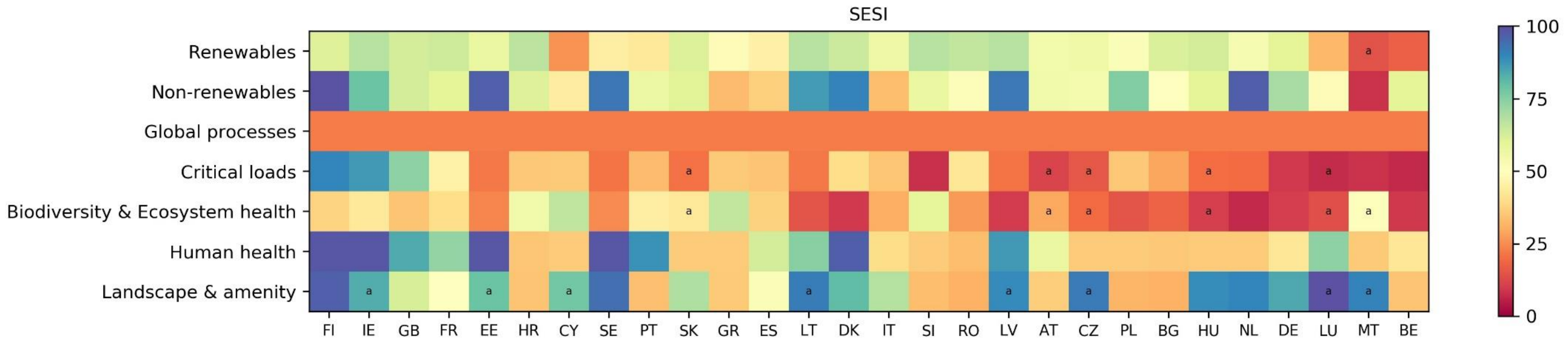
Figure 11: SESI scores by environmental function



The figure shows the scores of each country for the four environmental functions. Dark red indicates low scores, while dark blue indicates high scores. Countries are sorted by the total SESI score from higher to lower.

The label [a] in the heatmap indicates that one of the indicators assigned to the principle is blank because it does not apply to the country. These includes (1) indicators on coastal waters in AT, CZ, HU, LU and SK; (2) indicators on World Heritage sites in CY, CZ, EE, IE, LV, LT, LU and MT; (3) and indicators on the ecological status of freshwater systems and on forest resources in MT.

Figure 12: SESI scores by sustainability principle



The figure shows the scores of each country for seven sustainability principles. Dark red indicates low scores, while dark blue indicates high scores. Countries are sorted by the total index score from higher to lower.

Note: The labels in the y axis are equivalent to the following principles in Table 4. Renewables: renew renewable resources; Non-renewables: use non-renewables prudently; Global processes: prevent global warming; Critical loads: respect critical loads for ecosystems; Biodiversity & Ecosystem health: maintain biodiversity and ecosystem health; Human health: respect standards for human health; Landscape & amenity: conserve landscape and amenity.

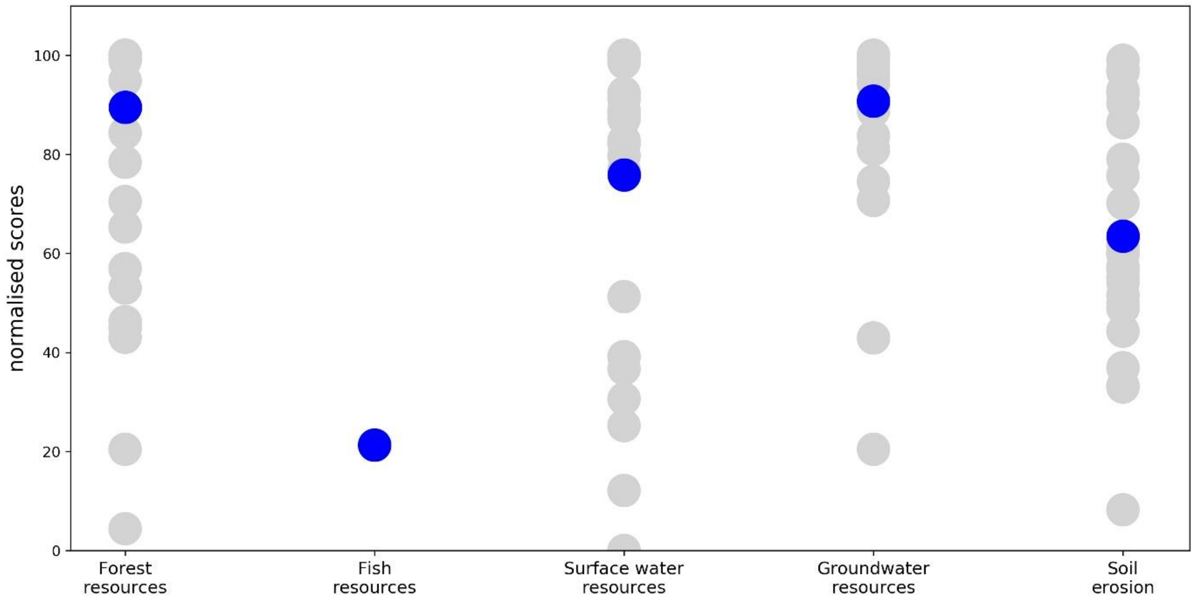
The label [a] in the heatmap indicates that one of the indicators assigned to the principle is blank because it does not apply to the country (e.g. coastal areas in inland countries).

The following subsections provide more details on how EU countries perform against environmental standards across the four main functions. As explained before (c.f. section 4.4), in most indicators, the normalised score shown in the next figures represents the % of an element (e.g. ecosystem, population) that meets the corresponding environmental standard. All the information on the indicators and their environmental standards is given in section 1 of Annex 1.

4.9.1. Source function

Indicators of source function relate to provision of natural resources and measure the sustainable exploitation of renewable and non-renewable natural capital assets. As shown in Figure 13, countries obtain high scores in the exploitation of most renewable resources. In the case of forest resources, Europe obtains a score of 90 with an estimated 73% wood utilization rate in 2015, which suggests that on average, its wood supply is not threatened over the long term. This is the situation in most countries. In fact, half the countries obtain a full score of 100, which indicates that their forest utilization rate is below the 70% standard, which improves the forest's potential for wood production, and the conditions it provides for biodiversity, health, recreation and other forest functions EEA (2017). Most countries report utilization rates between 70% and 100%, with only Belgium reporting exploitation rates close to 100%, which if sustained over time may result in unsustainable production and lead to younger forests, lower biomass pools, depleted soil nutrient stocks and a loss of other ecosystem functions (Schulze et al. 2012).

Figure 13: Normalised SES indicator scores for the source function



Blue dots indicate the performance of the European block in the different indicators. Grey dots, on the other hand, represent the performance of the individual countries.

The exploitation of freshwater resources also shows a wide distribution with nine countries reporting normalised scores of 100, which indicates that the Water Exploitation Index + (the ratio between the consumption of freshwater and renewable freshwater resources at a given point in time) was below the 20% standard in every quarter of the year 2015, thereby indicating no water stress as defined by Raskin et al. (1997). In Europe, 76% of the river basin area had not experienced seasonal water stress conditions. Seasonal water stress is more common in Southern European countries, where six countries obtain a score of 51 or lower. This shows that at least half of the river basin area in those countries was subject to water stress conditions at least in one quarter of 2015. In the islands of Cyprus and Malta, as well as in Luxembourg, all the river basins were subject to water stress conditions at least in one season during the year. Although water stress is more pronounced in the summer in Southern European countries, it is not a problem specific to that region or that season alone (Zal et al. 2017). Belgium and the Netherlands also score quite low, and in several Central and Eastern European countries, between 10 and 20% of the water basin area was subject to stress conditions throughout the year.

In Europe, 91% of the area covered by groundwater bodies was assessed as being in good quantitative status during the second reporting period of the Water Framework Directive, which started in 2015. A groundwater body is considered to be in good quantitative status when the available groundwater resource, which depends among others on its replenishment rate, is not exceeded by the long-term annual average rate of abstraction (EC 2009). In general, countries report high scores, with 22 countries reporting 85% of the groundwater area being in good quantitative status. This can, nonetheless, hide relevant spatial disparities where a large share of river basins is overexploited. This is for instance the case of Spain, Italy and the United Kingdom. In Spain, 93% of the groundwater area of the Tagus is overexploited, while in other relevant river basins such as Guadiana and Segura, more than half of the area is not exploited sustainably. In Italy, more than half of the groundwater resources in Padano, Appennino Meridionale and Sicilia do not meet the environmental standard. In the United Kingdom, most of the groundwater resources in England are overexploited (e.g. 78% in the Thames river basin). At the bottom of the country list, in Cyprus and Malta only 43% and 20% of their groundwater area meets the environmental standard. In these countries, more than half of their water needs are met through groundwater resources (EEA 2018b). In Europe, the main reasons behind the overexploitation of groundwater bodies relate to the water balance or lowered water table (75%), the deterioration of related surface waters (24%) and dependent terrestrial ecosystems (20%), and saline intrusion (9%) (EEA 2018b).

The exploitation of fish resources is the indicator related to renewable resources in which Europe performs the worst. In this indicator, the same score is assigned to all the countries, given that even though coastal countries have their own exclusive economic zone, the fisheries policy – and therefore fishing quotas – are determined at European level. Thus, only a single score is apparent in Figure 13. In 2017, out of the 228 commercial fish and shellfish stocks for which information was available, only 21% were in good environmental status [11]. In order to meet the environmental standard, both the mortality and reproductive capacity of fish stocks need to be consistent with the Maximum Sustainable Yield (EC 2010). The underlying data shows relevant regional differences. The situation is dramatic in the Mediterranean and Black Sea, since none of the fish stocks assessed meets the environmental standards. On the opposite end, in the Bay of Biscay & Iberia, the Celtic Seas and Greater North Sea, around half of the stocks assessed are in good environmental status.

Soil erosion is the only indicator covering non-renewable resources. In Europe, 63% of the territory had soil erosion rates above the environmental standard (1 t ha^{-1} yr^{-1}). Five

[11] The indicator score would increase to 49 if the environmental standard would be relaxed to consider only one criterion (mortality or reproductive capacity).

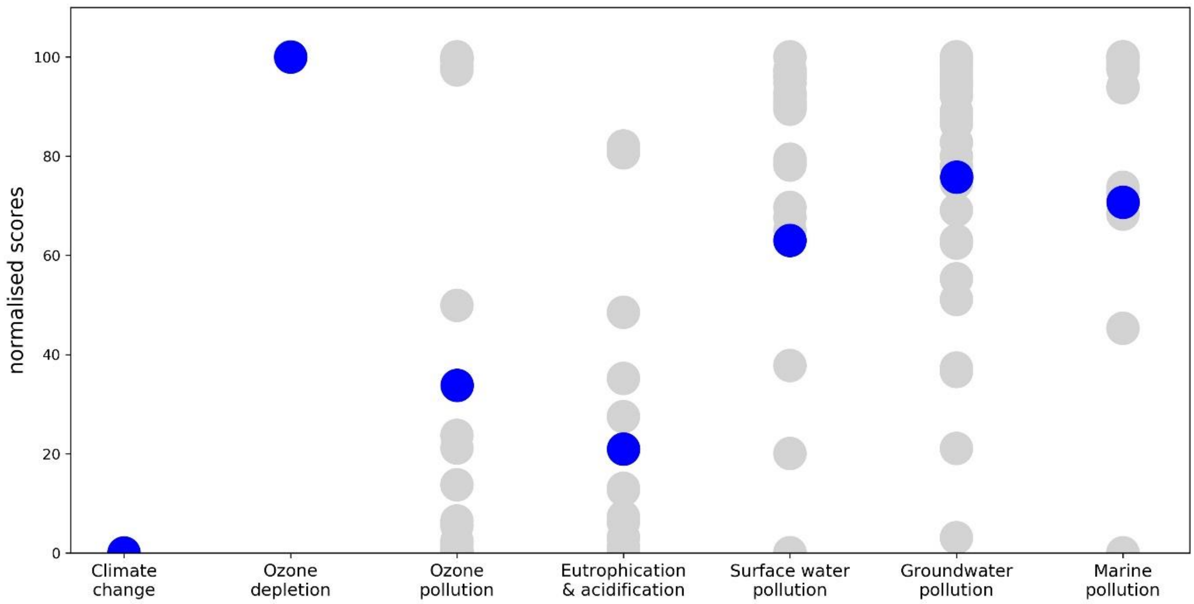
Mediterranean countries are the most affected countries with scores below 50. The effects of soil erosion are also apparent in mountainous regions. Scores above 85 are found in Scandinavian countries, the Baltic states and the Netherlands.

4.9.2. Sink function

Figure 14 shows the country scores for the indicators in the sink function. The sink function covers the disruption of Earth System processes, as well as the pollution of different types of ecosystems. Since Earth System process indicators are represented through pressures, the scores have been obtained by normalising the indicators with upper and lower bounds determined by environmental standards. In the case of ecosystem pollution indicators, the scores show the percentage of ecosystem area that meets relevant environmental standards that define their chemical status.

European countries obtain completely opposite scores in the climate change and ozone depletion indicators. CO_2 emissions of European countries were in the 3-16 t per capita range in 2018, with around 7 tonnes per capita for the European block. All these values are above the 0.5-2.5 tonnes per capita range that would lead to a global mean temperature change between 1.5-2ºC compared to the preindustrial period. In the case of consumption of ODS, Europe and the individual countries are assigned the same score because the EU reports its data in aggregate form. Nonetheless, most companies that report data are located in Germany, the United Kingdom, France and Italy (EEA 2019a). In 2019, Europe reported a consumption of 61 tonnes of ozone depleting potential, which leads to a normalised score close to 100.

Figure 14: Normalised SES indicator scores for the sink function



Blue dots indicate the performance of the European block in the different indicators. Grey dots, on the other hand, represent the performance of the individual countries.

The picture is more diverse when looking at the pollution in terrestrial, freshwater and marine ecosystems. There are two indicators used to characterise pollution in terrestrial ecosystems: ground-level ozone pollution in agricultural land and forests, and the eutrophication and acidification of terrestrial ecosystems. In both cases, less than 35% of the European area considered meets the environmental standards. In the case of ozone pollution, Scandinavian countries, Baltic states and the Anglo-Celtic isles get almost perfect scores (>95) in 2017. Ozone pollution is widespread in the remaining countries. In half of them, less than 5% of the agricultural land and forest area is below critical levels of ground-level ozone. In the case of acidification and eutrophication, around a third of the European natural and semi-natural terrestrial ecosystem area meets the environmental standards. The rest exceeds either the critical loads of acidification or eutrophication (or both). As in the case of ozone pollution, Finland and Sweden are the frontrunners, although in both cases around 20% of the relevant area does not meet the environmental standards. Except in the Anglo-Celtic isles, where between a third and half of the territory is in good condition, in the rest of the European territory the transgression of the critical loads of acidification and/or eutrophication is the norm. These two issues affect different parts of the continent though. Eutrophication of terrestrial ecosystems is widespread across Europe, although its severity differs considerably. As noted by Fagerli et al. (2020), the most severe exceedances are in the Po Valley in Italy, the Dutch-German-Danish border areas and in some parts of Spain close to the Mediterranean Sea. Acidification, on the other hand, is a less prevalent problem in Europe. In this case, the main hotspots can be found in the Netherlands, Belgium and parts of Germany.

Freshwater systems show a better picture than terrestrial ecosystems. Freshwater systems are assessed through indicators on the chemical status of rivers and groundwater systems. The chemical status of rivers is determined based on the compliance with the environmental quality standards (defined as concentration of pollutants in water) established in European legislation. These are intended to protect the most sensitive species from direct toxicity, including predators and humans via secondary poisoning (European Parliament and European Council 2008b). In Europe, 63% of the rivers (measured in length) met the environmental standards. Regional differences are very apparent in country performances. 13 countries (mostly Mediterranean countries, the Anglo-Celtic isles and some Eastern European countries report that more than 90% of their rivers were in good chemical status in 2015. On the other end, Central European countries such as Germany, Austria, Luxembourg, Belgium, as well as Slovenia and Sweden reported virtually every river failing to meet the environmental standard for at least one of the chemicals identified as priority substances in the European legislation. In the Netherlands and Denmark most rivers were also in poor chemical status. The main reason for failing to meet the environmental standards is the excessive presence of mercury and brominated diphenyl ethers in water bodies, which are considered ubiquitous, persistent, bioaccumulative and toxic substances (EEA 2018b). Part of this pollution relates to legacy sources, or in the case of mercury, natural sources as well. The results are very sensitive to the inclusion of ubiquitous, persistent, bioaccumulative and toxic substances in the analysis. When excluding them, the number of water bodies that meet the environmental standards increases significantly (EEA 2018b).

The chemical status of groundwater bodies depends on two factors: saline intrusion and concentration of pollutants. The latter needs to comply with the environmental standards set in European legislation and be consistent with achieving good status in dependent surface water ecosystems (EC 2009). In Europe, 76% of the area covered by groundwater bodies met the environmental standards. Low scores are obtained in areas with intensive agricultural activity or with past or present heavy industry (EEA 2018b). In fact, nitrates and pesticides account for most of the groundwater bodies in poor chemical status. 15

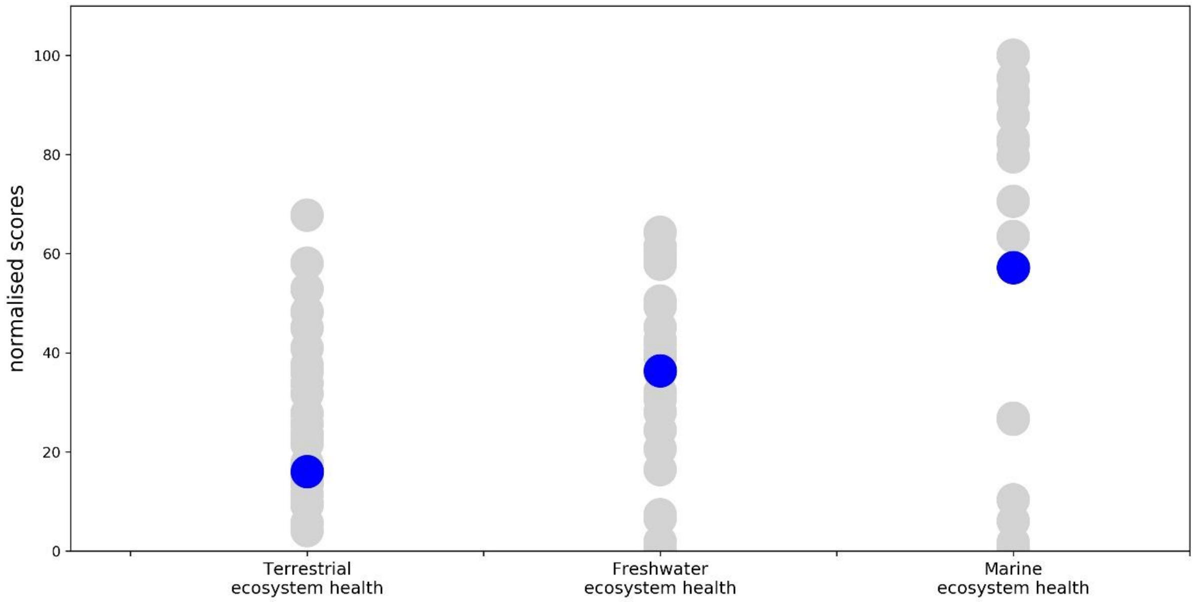
countries obtain scores above 85, while in countries such as Malta, Luxembourg, Belgium and the Czech Republic less than 40% of the groundwater area is in good chemical status.

The chemical status of coastal systems is used as a proxy to assess marine ecosystems, since European countries have not reported data on the latter yet. Environmental standards that characterise good chemical status are the same as in the case of surface water bodies. In 2015, 71% of the European coastal area was in good chemical status. In all the Baltic states, and in countries with high marine transport activity such as Germany, Belgium, the Netherlands and Sweden the coastal area that met the environmental standard was below 5%. Denmark, which also has a significant marine transport activity, obtained a score of 73. In contrast, all the Mediterranean countries except Italy, Romania, the Anglo-Celtic isles and Finland scored above 94. This indicator does not apply to landlocked countries such as Austria, Czech Republic, Hungary, Luxembourg and Slovakia.

4.9.3. Life support functions

Life support functions are characterised through ecosystem condition indicators that consider different criteria to assess ecosystem health in terrestrial, freshwater and marine ecosystems. The results of European countries are shown in Figure 15, although caution is advised when comparing country performances (see indicator fiches).

Figure 15: Normalised SES indicator scores for the life support function



Blue dots indicate the performance of the European block in the different indicators. Grey dots, on the other hand, represent the performance of the individual countries.

Ecosystem health in terrestrial ecosystems is determined based on the criteria outlined in the Habitats Directive. In Europe, 16% of the terrestrial ecosystems assessed have been classified as being in good conservation status. All the terrestrial ecosystem types perform poorly, but differences between classes are apparent. For instance, dune habitats, and bogs, mires and fens tend to be in worse status than rocky mountains (EEA 2020b). Relevant differences are also apparent at country level. Only three countries have more than half of their terrestrial ecosystems in good conservation status, while in 11 countries less than a fifth of the ecosystems meet the environmental standards. For an ecosystem to be in favourable conservation status – and therefore comply with the environmental standard – it needs to meet certain conditions related to its natural range, structure and functions, and biodiversity (Röschel et al. 2020). The ecosystem's structure and functions, as well as its future prospects tend to be the main reason for failing to achieve the environmental standards. In contrast, the habitat range is commonly in better status (EEA 2020b).

Freshwater ecosystems are in better shape than terrestrial ecosystems, although their condition is far from being sustainable. In Europe, 36% of rivers (measured in length) are in good condition considering the biological, chemical and hydromorphological criteria outlined in the Water Framework Directive. Country scores go from 64% in Finland to 0% in the Netherlands, where none of the rivers meet the environmental standards. In this vein, only eight countries have more than 50% of their rivers in good ecological status. The northern parts of the Scandinavian region and the United Kingdom, as well as Estonia, Romania, Slovakia as some parts of the Mediterranean countries have the highest proportion of river bodies that meet the environmental standards (EEA 2018b). Physical alternations to water bodies and structures such as dams, barriers and locks are among the main hydromorphological pressures leading to changes in habitats, which ultimately results in failure to achieve good ecological status in many freshwater bodies. Diffuse pollution (mainly from agriculture) as well as atmospheric deposition (mainly of mercury) are also relevant pressures on freshwater ecosystems that lead to nutrient enrichment and chemical pollution. Point source pollution such as that originated through urban wastewater treatment is less widespread, although by no mean negligible (EEA 2018b).

In the absence of data on the health of marine ecosystems, the ecological status of coastal waters has been used as a proxy. In Europe, 57% of the coastal area was in good ecological status. Country scores differ considerably with five countries having more than 90% of the coastal area meeting the environmental standards, and seven where all the coastal area fails to achieve them. The picture is particularly grim in the Baltic Sea, where the vast majority of coastal waters are not in good ecological status.

4.9.4. Human health and welfare function

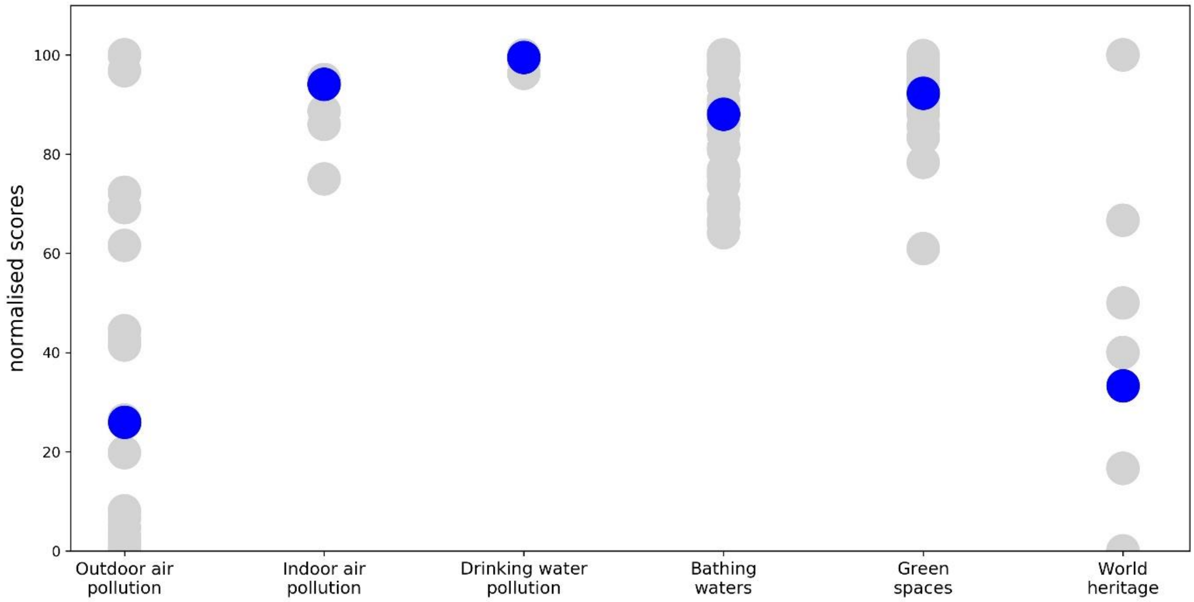
The indicators on human health and welfare functions can be split into human health and other welfare. The former capture how humans are affected by different pollution sources, while the latter consider whether other forms of welfare are threatened.

Human health is represented through outdoor and indoor air pollution and drinking water pollution indicators. In 2017, only 26% of the European population lived in areas where the annual average $PM_{2.5}$ concentration did not exceed the guideline value proposed by the World Health Organisation (WHO 2005). The air quality was better in Northern countries such as Finland, Sweden, Ireland and Denmark, where almost all the citizens were not exposed to $PM_{2.5}$ concentrations above the environmental standard. In 15 countries, more than 95% of the population lived in areas that did not met the environmental standard. Eastern European countries performed the worst in this indicator, not only in the extent of the transgression, but also on the severity. For instance, in Bulgaria, Croatia, Hungary and Poland around a third of the population was exposed to air pollution levels that were twice as high as the environmental standard. In Romania and Slovakia, the percentage exposed to twice the environmental standard was above 60%.

Regarding indoor air pollution, Europe performs quite well compared to less industrialised countries where the use of solid fuels for cooking and heating leads to important health impacts. In Europe, the majority of the population had access to clean cooking fuels and technologies. The use of solid fuels such as wood, crop residues and coal leads to exposure to particulate matter that is several times higher than the daily values proposed by the World Health Organisation (WHO 2018a). Thus, it represents the most severe forms of indoor air pollution, although other forms are not covered in this assessment. In 2018, Europe was assigned a score of 94, although this is likely higher given that in the original source, the maximum percentage of population with access to clean fuels within the household is denoted as >95. Out of the 28 countries considered, 25 had the maximum score.

European countries also perform very well in relation to drinking water pollution. In total, an average of more than 99% of the country samples analysed in 2013 met the environmental standards related to microbiological, chemical and other parameters. Except Malta, which scored 96 points, the remaining countries had normalised scores higher than 99.

Figure 16: Normalised SES indicator scores for the human health and welfare function



Blue dots indicate the performance of the European block in the different indicators. Grey dots, on the other hand, represent the performance of the individual countries.

Other welfare functions are represented through bathing water quality, access to green spaces, and relevant World Heritage site indicators. Out of the 21,500 bathing sites assessed in Europe in 2019, 88% were classified as excellent considering the environmental standards for *E. Coli* and intestinal enterococci. Generally, coastal bathing sites were in better shape than inland sites due to the influence of short-term pollution in small lakes and ponds, and in low-flow rivers due to heavy rains during the summer period (EEA 2020c). Country-wise, six European countries (Luxembourg, Cyprus, Austria, Greece, Malta and Croatia) got normalised scores of 95 or higher. Of course, the number of bathing sites in each country differs considerably. For instance, Luxembourg reports data on reports data on 12 bathing sites, while Greece does the same on almost 1,600. In most countries, 70 to 95% of their bathing sites meet the environmental standards. Slovakia (69), United Kingdom (67), Bulgaria (66) and Estonia (64) sat at the bottom of the list.

The indicator on green spaces measures the percentage of urban population that have green areas such as parks and forests within a 10-minute walking distance. In European countries, 92% of the urban population (ca. 200 million inhabitants) met the environmental standard of having access to nearby green spaces in 2012. The normalised score was above 90 in 21 out of the 28 countries considered. Five Mediterranean countries and Romania were the worst performers. These countries did not only have the highest proportion of population without access to nearby green spaces, but also had the lowest median surface area of accessible green urban spaces (Poelman 2018).

The last indicator related to other forms of welfare provision addresses the natural and mixed World Heritage sites. In 2020, only a third of the 36 sites that overlap with the 28 European countries considered were found to be in good conservation status based on the current state and trend of the values for which the site was designated, the threats affecting those values and the effectiveness of the conservation measures put in place to conserve those values. In this context, the environmental standard is defined through the 'good' status, which indicates that the site's values are in good condition and under the existing conservation measures they are likely to be maintained in the future (Osipova et al. 2014). In the remaining cases, there were either a few concerns over their status or a significant concern. None of the sites were in critical conditions. Countries reported very different performances. While all the relevant sites in Finland, Sweden, the Netherlands and Hungary were in good condition, none were in nine other countries. Further, eight countries did not have any natural or mixed World Heritage sites within their borders.

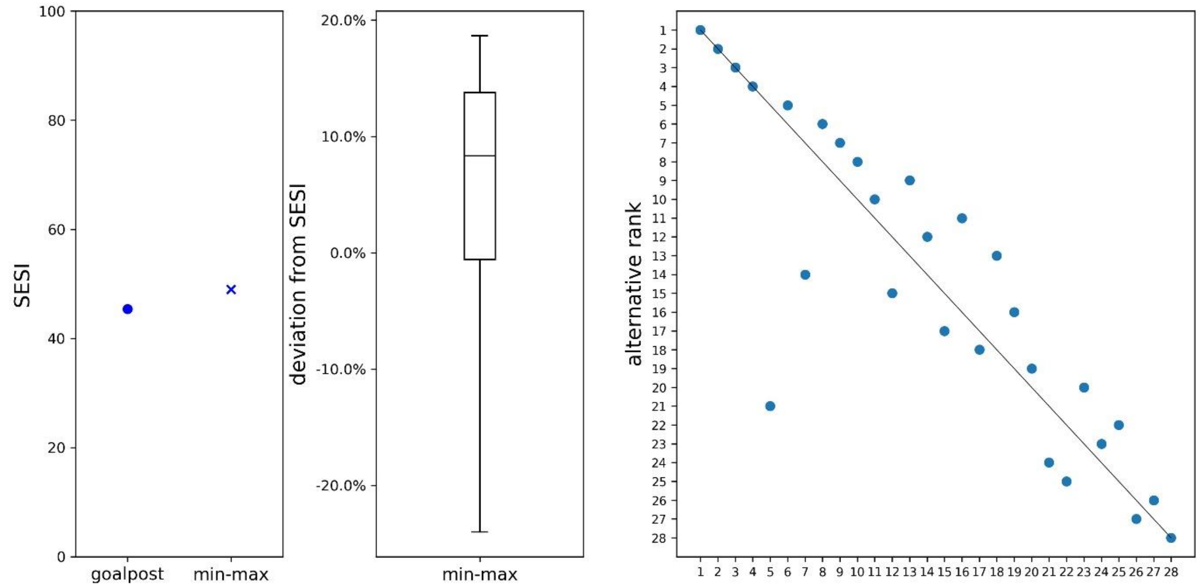
4.9.5. Uncertainty analysis

The uncertainty analysis presented here has been designed to test how the results are affected by assumptions related to relevant aspects of the theoretical framework. The uncertainty associated with the normalisation method, set of weights, treatment of zeros and small values, as well as the aggregation method is described in the following sections.

4.9.5.1. Normalisation method

Environmental sustainability demands measuring absolute performance, which can be interpreted as calculating normalised country scores in relation to environmental standards. SESI uses the goalpost method for such task. Figure 17 shows the effects of measuring relative performance instead of absolute performance. Relative performance is represented through the min-max method where country scores are a function of the best and worst performers. As shown in the figure, the index score at European level changes from 45.20 to 49.01 if instead of measuring absolute performance, relative performance is measured. At the country level, relevant deviations are seen in both directions when using the min-max method to normalise the underlying indicators (median=8.86%). When it comes to country rankings, the choice of the normalisation method also affects the results. 17 countries are in the [-2, 2] range, with some outliers, who lose 16 and seven positions when considering relative performance.

Figure 17: Uncertainty associated with the normalisation method of SESI at index level

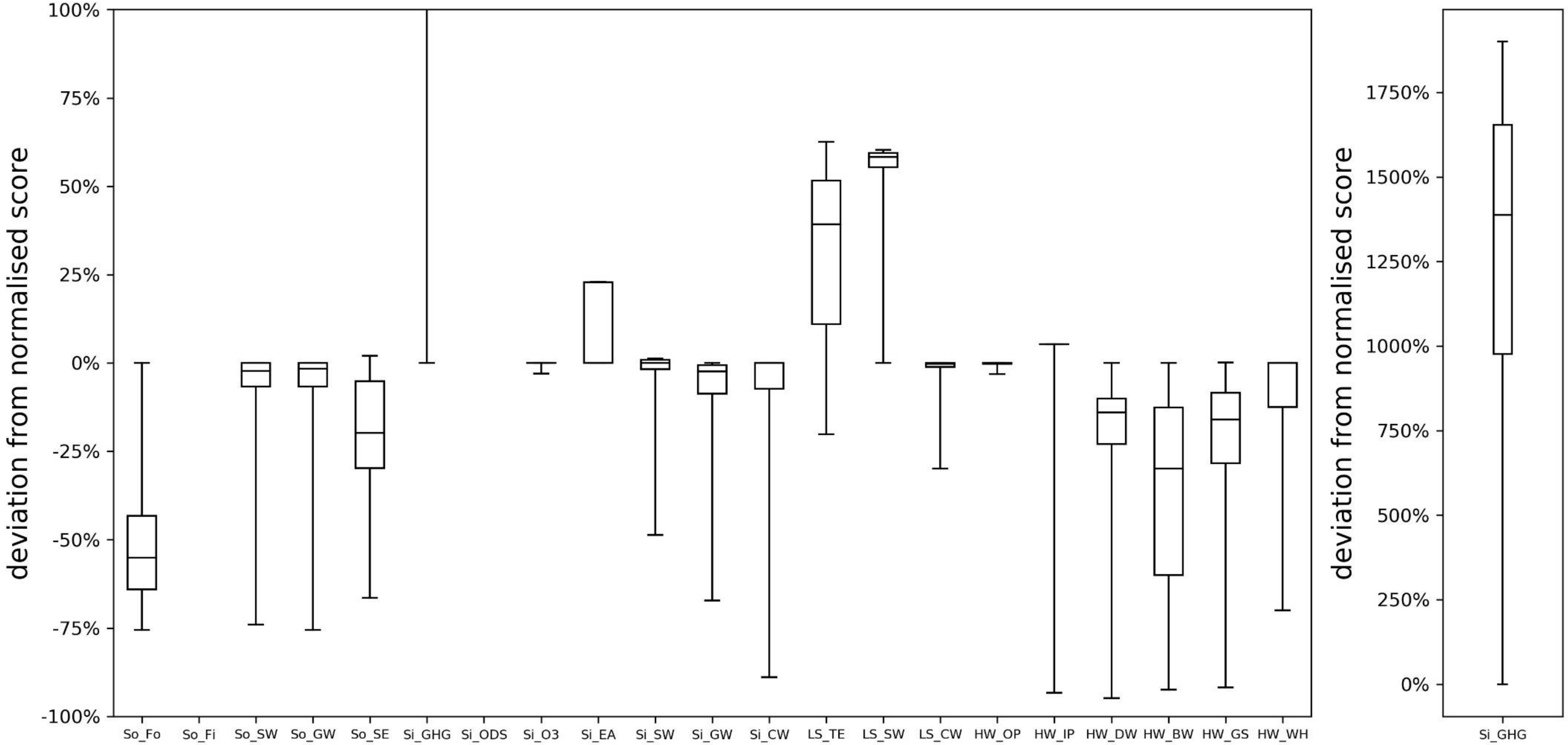


The figure on the left shows the index score for Europe with the default (goalpost) and min-max method. The figure in the centre shows the distribution of the differences between min-max and the default method at country level. The figure on the right shows the rank comparison between the default (goalpost) and min-max method at country level.

The upper and lower edges of the rectangle in the boxplot represent the 75th and 25th percentiles, while the top and bottom markers represent the maximum and minimum values.

The effect of the normalisation method in each indicator differs greatly as shown in Figure 18, although some patterns emerge. For instance, for indicators where European countries consistently perform well in absolute terms (e.g. drinking water quality, indoor air pollution and quality of bathing waters), the normalised scores using the min-max values are much lower given that the lowest performer is a good performer when using the goalpost normalisation method. The opposite holds true for indicators where countries perform poorly in absolute terms (e.g. conservation status of terrestrial ecosystems, ecological status of freshwater ecosystems). In this case, the normalised scores using the min-max method are higher than the normalised scores using the goalpost method. CO_2 emissions represent a special case in this group, since all the countries have a normalised score of zero, which is converted into five after treating zeros and small values. Given that the original CO_2 emissions per capita varied between European countries, the maximum normalised score using the min-max method is 100, a 2000% higher than the values assigned with the goalpost method. In the figure, the indicators on fish stocks and consumption of ODS do not have valid data because the same not normalised score has been assigned to all the countries. The variations in other indicators depend on the maximum and minimum scores of the country sample.

Figure 18: Uncertainty associated with the normalisation method of SESI at indicator level



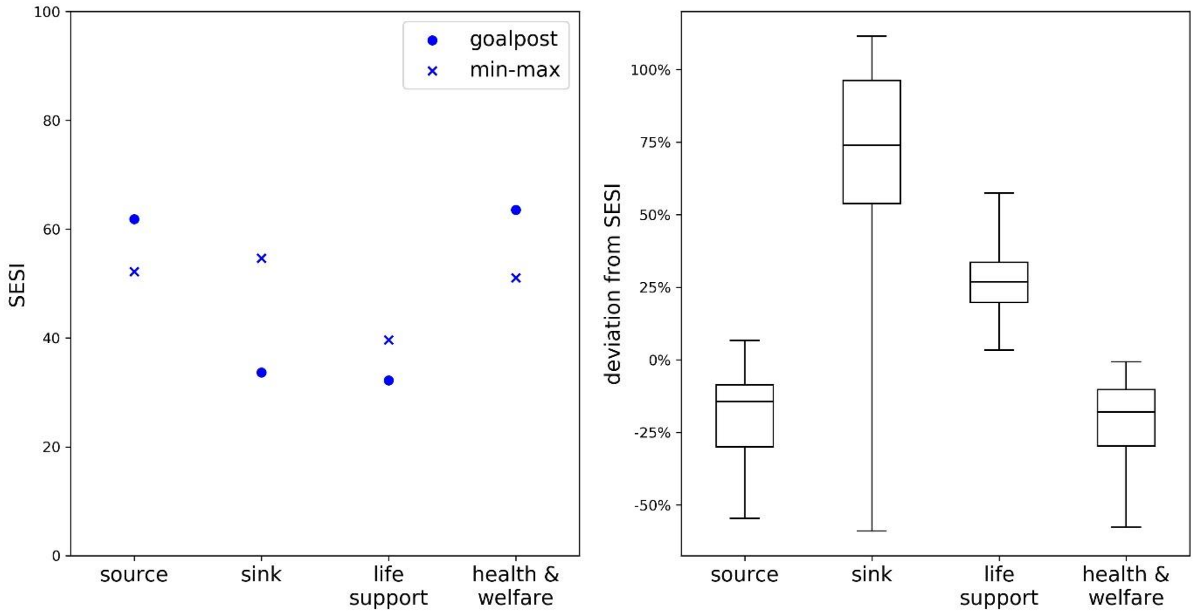
The figure on the left shows the information for all the indicators, expect for CO_2 emissions, which is shown on the right.

The upper and lower edges of the rectangle in the boxplot represent the 75th and 25th percentiles, while the top and bottom markers represent the maximum and minimum values. The normalised values have been treated to avoid zeros as described in section 4.6.

So_Fo: Forest utilization; So_Fi: Fish stocks; Surface water scarcity, So_GW: Groundwater scarcity; So_SE: Soil erosion; Si_GHG: CO_2 emissions; Si_ODS: Consumption of ODS; Si_O3: Ozone pollution in terrestrial ecosystems; Si_EA: Eutrophication and acidification in terrestrial ecosystems; Si_SW: Chemical pollution in surface waters; Si_GW: Chemical pollution in groundwater; Si_CW: Chemical pollution in coastal waters; LS_TE: Ecological health of terrestrial ecosystems; LS_SW: : Ecological health of surface waters; LS_CW: Ecological health in coastal waters; HW_OP: Outdoor air pollution; HW_IP: Indoor air pollution; HW_DW: Drinking water quality; HW_BW: Quality of bathing waters; HW_GS: Proximity to green spaces; HW_WH: Conservation of World Heritage sites.

Given that the normalisation method has different effects in each indicator, the function scores are also affected differently. As shown in Figure 19, the scores for the source, and human health and welfare functions tend to be lower with the min-max normalisation, while the opposite holds for sink and life support functions. This holds for the scores at European level, where the source, and human health and welfare scores decrease from 61.82 and 63.53 with the goalpost normalisation to 52.20 and 51.04 with the min-max normalisation. In the case of sink and life support functions, the scores increase from 33.01 and 32.21 to 54.64 and 39.64. A similar effect can be seen for country scores, although exceptions apply.

Figure 19: Uncertainty associated with the normalisation method of SESI at function level



The figure on the left shows the function scores for Europe with the default (goalpost) and min-max method. The figure in the centre shows the distribution of the differences between min-max and the default method at country level for each function.

The upper and lower edges of the rectangle in the boxplot represent the 75th and 25th percentiles, while the top and bottom markers represent the maximum and minimum values.

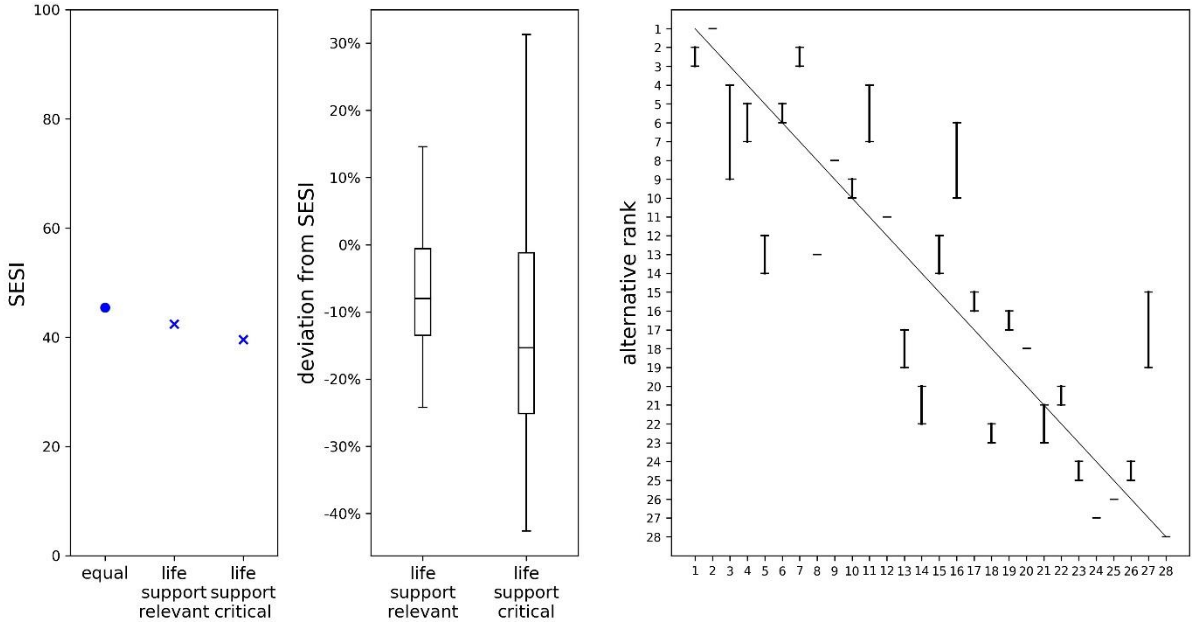
In summary, measuring absolute or relative performance yields significantly different results in the case of SESI, its function scores and the indicators. This has relevant implications for the interpretation of the results, since only the goalpost method is aligned with the concept of environmental sustainability and science-based standards adopted in the ESGAP framework.

4.9.5.2. Weights

Different sets of weights have been tested by assigning more implicit relevance to the life support function. As shown in Figure 20, changing the sets of weights leads to lower index values at European level and for most countries. The more weight is assigned to the life support function (0.4 under 'life support relevant', and 0.7 under 'life support critical'), the lower the overall index score. This is the result of assigning more weight to a dimension in which countries tend to perform worse than in others. The right side of the figure shows that country ranks are also considerably affected.

Equal weights have been chosen as the default option due to the absence of an internationally agreed weighting method. It should be noted that the uncertainty of the weighting method used to weight functions can be mitigated by displaying the function scores alongside the index scores as concluded from the statistical coherence analysis. Nonetheless, this does not address uncertainty related to the weighting of indicators, as opposed to functions. This issue has not been explored here because assigning a set of weights to the indicators would require considering the context of each country, which would ultimately hamper the comparability of the index results.

Figure 20: Uncertainty associated with the weighting method of SESI at index level



The figure on the left shows the index score at European level with the default (equal) weights, and two additional sets of weights. The figure in the centre shows the distribution of the differences between the default and the other two sets of weights at country level. The figure on the right compares the ranks obtained with the default and the other two sets of weights.

The x axis in the first two figures shows the different sets of weights chosen for the source, sink, life support, and human health and welfare functions. 'Equal' assigns 0.25-0.25-0.25-0.25 to the functions. 'Life support relevant' assigns 0.20-0.20-0.40-0.20, while 'life support critical' assigns 0.10-0.10-0.70-0.10.

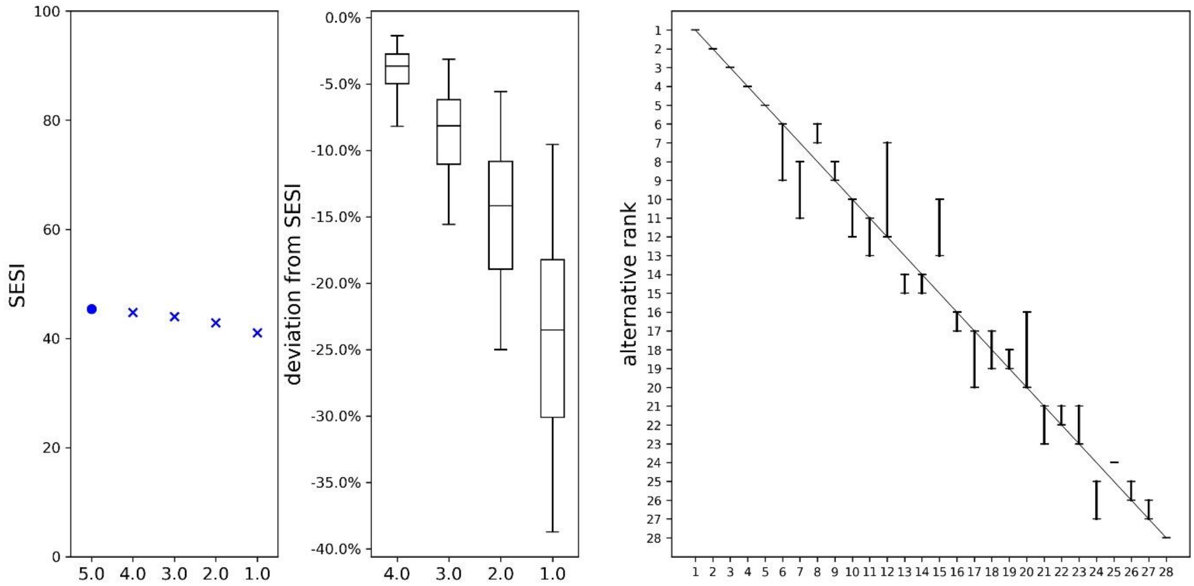
The upper and lower edges of the rectangle in the boxplot represent the 75th and 25th percentiles, while the top and bottom markers represent the maximum and minimum values.

4.9.5.3. Treatment of zeros and small values

As argued in Box 3, indices that use geometric mean to aggregate information usually treat zero and small values in order to avoid the mean collapsing to zero when small values are present. After comparing the country ranks obtained using the Borda and Copeland methods with those obtained after testing different values to treat zeros and small values (c.f. section 4.6), the value five was chosen.

The results in Figure 21 suggest that the index score at European level is barely affected by the choice of the value to treat zeros and small values. At the country level, changes can be substantial. The lower the value, the lower the index scores. The rankings, on the other hand, seem to be relatively insensitive, although with a few exceptions. The effect of the value chosen in the index scores depends on the number of indicators that are treated. For this reason, the effects are more evident for the sink and life support function scores, since more countries tend to have very low scores in the underlying indicators.

Figure 21: Uncertainty associated with the treatment of zeros and small values in SESI at index level



The figure on the left shows the index score at European level with the default value used to treat zeros and small values (five), and four alternatives. The figure in the centre shows the distribution of the differences between the country scores obtained using the default value and the alternatives tested. The figure on the right compares the ranks obtained using the default method and the alternatives tested.

The x axis in figures a and b show the different values chosen to treat zeros and small values.

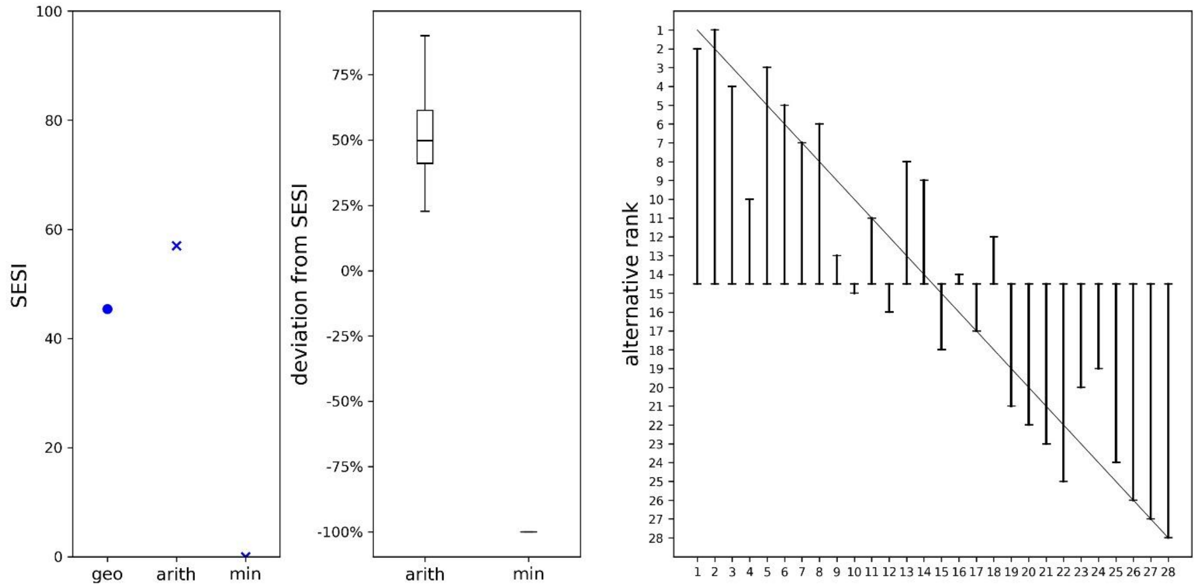
The upper and lower edges of the rectangle in the boxplot represent the 75^{th} and 25^{th} percentiles, while the top and bottom markers represent the maximum and minimum values.

4.9.5.4. Aggregation method

The aggregation method used significantly affects SESI scores both at European and at country level as shown in Figure 22. The use of the arithmetic mean increases the score of Europe from 45.20 to 56.55. At country level, the median increase is 49%. The effect is most obvious in the score of the sink functions, since countries perform particularly poorly in several of the indicators in this dimension.

Using the minimum indicator value, which assumes no substitution capacity at all between the functions of natural capital, completely disrupts the results, since all the countries have a normalised value of zero for the indicator on CO_2 emissions. As a result, the index value will be zero for all the countries and all the countries will have the same position in the rankings.

Figure 22: Uncertainty associated with the aggregation method of SESI at index level



The figure on the left shows the index score at European level with the default (geometric mean) and alternative methods (arithmetic mean, minimum value). The figure in the centre shows the distribution of the differences between the country scores obtained using different aggregation methods. The figure on the right compares the ranks obtained with different aggregation methods.

The x axis in figures a and b show the different aggregation methods tested. Geo: geometric mean; arith: arithmetic mean; min: minimum value.

The upper and lower edges of the rectangle in the boxplot represent the 75^{th} and 25^{th} percentiles, while the top and bottom markers represent the maximum and minimum values.

4.10. Discussion

4.10.1. A novel index of strong environmental sustainability

SESI is a single metric that represents the extent to which countries comply with environmental standards intended to represent the conditions under which the functioning of natural capital is not threatened in the long term. The index comprises indicators that are aggregated across different layers: from indicators to topics, from topics to sustainability principles, from principles to the main function of natural capital (source, sink, life support, and human health and welfare), and from the latter to a single index. Each of these indicators measures absolute performance against a science-based environmental standard, thereby showing whether specific functions of natural capital are potentially compromised.

The selection of indicators has been done based on the relevance of the indicator, its statistical and methodological soundness, and data quality. This allowed shortening the initial list of 30 candidates into the 21 indicators that form the SESI. The relevance criterion required the indicator to (a) be related to the functions of natural capital, (b) to have a science-based reference value against which performance can be measured, and (c) to be defined at the national level. Relevance assesses consistency with the definition of environmental sustainability adopted in the ESGAP framework. Statistical and methodological soundness, as well as data quality considered more generic criteria that can be applied to any other index. While the 21 indicators cover quite a lot of ground, there are some topics such as extraction of raw materials, organic soil matter, marine systems or several aspects of human welfare that are not covered in this version of the index. This is the result of a lack of environmental standards or data for the relevant indicators. As knowledge on environmental standards and data availability improves, the list of indicators should be revised.

SESI covers a space in sustainability science that none of the most widely known environmental (sustainability) indices covers. The main novelty of the index is the use of science-based environmental standards to measure the absolute environmental sustainability performance of countries. Other indices such as the Environmental Performance Index or the environmental dimension of the SDG indicators either measure absolute performance against policy or international targets or relative performance against frontrunners (Lafortune et al. 2018; Yale University 2018; Eurostat 2019b; OECD 2019). Although some of the targets used are aligned with environmental standards, this is not the general rule. The information contained in these alternative indices is useful in many ways, but these metrics do not allow assessing the environmental sustainability of nations from a strong sustainability perspective.

Furthermore, it is worth noting that SESI has been built following the guidance provided by the most comprehensive manual on composite indicators (OECD and JRC 2008; JRC 2019), and therefore contains additional complementary analyses related to its conceptual and statistical soundness (see section 4 in Annex 1), which is already a distinctive feature compared to other metrics that tend to focus on the main results (Kwatra et al. 2020).

4.10.2. Are European countries environmentally sustainable?

The results suggest that the functioning of different elements of natural capital is impaired as a result of excessive environmental degradation in Europe. Most European countries obtain index scores below 50, including as the European block, which scores 47 points. In this context, it is important to bear in mind that only a score of 100 reflects compliance with the environmental standards of each of the 21 indicators selected to represent relevant environmental functions of natural capital. Even in the case of the highest scoring country Finland, the gap between the current and sustainable conditions is of 40 points.

Performance across environmental functions is quite uneven, with those related to environmental integrity being the most affected. In the sink function, countries perform very poorly with regard to CO_2 emissions and the chemical pollution of ecosystems, especially terrestrial ecosystems. Scores are also very low in the life support function, arguably the most important function of all, which covers biodiversity and ecosystem health. Functions associated with the provision of resources seem to be in better shape than those associated with the neutralisation of waste and life support. One can only hypothesise if the fact that biotic and abiotic resources have a market value can partially explain this pattern, which does not hold in every country. An exception in the source function are fish stocks, which are consistently overexploited across countries. This could be related to open access and free riding attitudes.

Countries tend to obtain relatively high scores when health standards are on the line as in the case of drinking water and indoor air pollution. Outdoor air pollution is an exception, arguably because the policy targets set are more permissive than the guideline values proposed by the World Health Organisation. When it comes to the amenity function, countries tend to have high scores in relation to bathing sites and access to green spaces, while with World Heritage sites, performance is very uneven with many countries not having any natural site within their territory.

The interpretation of the results needs qualifications on several grounds. First, the index provides a snapshot of whether countries meet science-based environmental standards from a territorial perspective across a variety of environmental and resource issues. While doing so, the indicators that form the index represent the extent to which environmental standards have been transgressed, but do not capture the severity of the transgression or the consequences of transgressing tipping points. For instance, the outdoor air pollution indicator represents the percentage of the population that is exposed to $PM_{2.5}$ concentrations higher than the guideline values proposed by the World Health Organisation. In theory, it would be possible for two countries to have the same normalised score (e.g. 75), while in the first country a quarter of the population is exposed to air pollution levels slightly above the environmental standards, while in the second a quarter of the population is exposed to air pollution levels that are several times higher than the environmental standard. In this case, severity could be represented through indicators on health impacts (e.g. the disability-adjusted life-years). In order to capture this dimension, in the future, the narrative developed through SESI should be complemented with severity indicators.

Second, the indicators that form the index adopt a territorial perspective, as opposed to the consumption perspective that is characteristic in environmental footprint indicators. SESI seeks foremost to be useful for policy making, and therefore is restricted to the elements of natural capital that can be most easily influenced by policy makers. Nonetheless, consumption-based indicators can provide a complementary perspective to the results, although the often lack the spatial dimension present in many of the indicators used here.

Third, the environmental standards used to characterise environmental sustainability have either been taken from the scientific literature or from relevant environmental legislation informed by expert input. Nonetheless, standards can refer to acceptable health risks, acceptable environmental impacts, precautionary expert guesses or safe distance from tipping points, and therefore, their meaning varies. Thus, the level of consensus around the standards chosen differs considerably. A key commonality of all the standards is that their transgression highlights a potential problem that demands policy attention. As the knowledge base improves, existing environmental standards might change, or new ones might be formulated. Likewise, it is important to bear in mind that potential trade-offs might arise when trying to meet environmental standards. For instance, the reduction of CO_2 emissions through bioenergy and carbon capture and storage would have negative impact on terrestrial habits (Heck et al. 2018a). Thus, interventions intended to address the environmental and resource issue covered in SESI should consider the potential consequences they might have in other areas.

Lastly, SESI provides a snapshot perspective on countries' environmental sustainability, and therefore fails to reflect whether progress towards the standard is being made over time. SESPI in chapter 5 is intended to fulfil this role.

4.10.3. Choices in the construction of the index matter

Indices have the potential to help make sense of complex systems through numbers. Nonetheless, the big picture they intend to show can be unintendedly distorted or even manipulated if the choices made during the construction of the index are not clear or properly justified (Greco et al. 2019). SESI is not exempt from such risk and therefore, its construction has been guided by the most comprehensive manual on composite indicators (OECD and JRC 2008). The computation of SESI required several methodological choices to be made in relation to data treatment, normalisation, weighting, and aggregation. When possible, key choices related to the indicator selection, normalisation or the aggregation have been aligned with the key features of the ESGAP framework, which reflects a more accurate and restrictive vision of the concept of strong sustainability as opposed to other indices such as the SDG Index or the Environmental Performance Index. These features have been summarised in Table 28.

The goal of the uncertainty analysis undertaken was to understand how choices in the construction of the index affect the results and the narrative developed based on them. Thus, different choices related to the normalisation, weighting, treatment of zeros and small values, and the aggregation were tested separately. The results show that the index and function scores are affected by these choices. Measuring the absolute performance of countries, which depends on the environmental standards, as opposed to measuring the relative performance of countries, which depends on the frontrunners and laggards leads to lower index scores, although functions and indicators are affected differently. For instance, measuring absolute performance for drinking water quality shows that virtually every European country complies with the environmental standards drawn in European legislation. When measuring relative performance, countries depend on the sample distribution and therefore even if a country performs well, it might get a low score if it is at the bottom of the distribution. This is the case of Malta, which is assigned the minimum score even if 96% of the samples analysed met the environmental standards. The latter clearly overestimates the real health risk drinking water quality poses in the country. The opposite holds for the conservation status of terrestrial ecosystems, where the best performer is Romania with 68% of the ecosystem assessments meeting the environmental standards. When measuring relative performance, this country is assigned a score of 100 for being the best performer, yet the data shows that 32% of the ecosystem assessments were not in good conservation status. When measuring relative performance, Romania would score 100 points as long as it was slightly above the second-best performer. This would be true in the 59%-100% compliance range. Thus, measuring relative instead of absolute performance through normalised scores would lead to biased messages in the context of environmental sustainability. After all, in many instances, best performances are not aligned with environmental sustainability conditions.

With regard to the substitution capacity between the functions of natural capital, the SESI uses the geometric mean with treatment of zeros and small values to represent a limited capacity in line with the strong sustainability discourse. Assuming full substitutability through aggregating with arithmetic means or no substitutability with the adoption of the minimum normalised score of any indicator as final index score significantly impacts the results. The use of the arithmetic mean leads to higher scores, especially in the functions in which countries perform worse. This makes it more challenging to identify which functions of natural capital are threatened if the low scores in the underlying indicators are linearly compensated by high scores. On the opposite end, when assuming no substitution capacity between functions, only the information on the worst performance is aggregated, which ultimately limits the usefulness of the index because it omits the information contained in all the other topics covered by the indicators. How zero and small values are treated when using the geometric mean has a less pronounced effect in the overall scores, but this can be partially mitigated by showing the function scores alongside the index scores when presenting the results. This recommendation also arises from the statistical coherence analysis.

The weighting method remains the most controversial choice in the construction of SESI. Equal weights have been assigned to all the indicators and (sub)dimensions, including functions. Indicator weights could be set based on the natural endowments of each country, but this would hinder the comparability of the results. At the level of function, the life support function has been identified as being more relevant than source, sink, and human health and welfare functions, but because of the lack of a generally agreed method to weight each function, equal weights have been used. The uncertainty analysis has tested different sets of weight at function level and the results show that their effect is by no mean negligible. Specifically, increasing the weights of sink or life support functions would generally lead to lower scores. As in the case of normalisation, showing the function scores alongside the index scores would minimise this effect. In any case, the issue of weighting remains unresolved in this version of the index and should be revisited in the future.

All in all, the uncertainty analysis has shown that the choices made during the construction of SESI are not trivial and therefore need to be aligned with the theoretical framework. After all, measuring absolute or relative performance, or assumptions about the substitution capacity between the functions of natural capital not only have an impact on the results, but also on the narrative built from them.

4.11. Conclusions

It is remarkable that countries still lack meaningful metrics that allow them to measure their environmental sustainability performance from a strong sustainability perspective. SESI is based on the ESGAP framework, which builds on key concepts such as strong sustainability, critical natural capital, and science-based standards of environmental sustainability. The limited substitution capacity between different types of capital and between the different functions provided by natural capital, and the notion that some elements of natural capital provide irreplaceable functions are much more closely aligned with the biophysical reality that governs the natural system and the socioeconomic systems embedded within, than the concept of weak sustainability, which assumes that the loss of nature can be fully compensated by increases in manufactured, human or social capital. For these reasons, metrics of weak sustainability can be misleading and lead to poor decision making.

Although this first version of SESI can only be considered a proof of concept, it can provide policy-relevant information by helping countries navigate the environmental sustainability agenda beyond single issues and providing scores that allow comparisons and benchmarking across countries. In this context, SESI provides a snapshot of the absolute performance of countries against environmental standards intended to represent whether the capacity of natural capital to provide ecosystem services is compromised. As a result, SESI provides a different perspective on the environmental sustainability of nations compared to most environmental indices and indicator systems that tend to measure the performance of countries against their peers or against policy targets, rather than science-based reference values (c.f. chapter 1).

At the outermost ring of Figure 6, the set of 21 indicators show the extent to which science-based environmental standards are met. Although there might be some overlaps with policy targets, the environmental standards adopted are meant to reflect the scientific understanding of good environmental quality. The resulting index is expected to differ from a potential policy gap index that could measure the gap between the current performance and existing environmental policy targets. The magnitude of the difference would depend on the extent to which environmental targets are aligned with science-based environmental targets.

At higher levels, SESI and the sub-indices for environmental functions (source, sink, life support, and health and human welfare) could be used as headline indicators when monitoring towards sustainable development at country level, thereby complementing the narratives around social and economic welfare. SESI could also feature in other broader development policies such as the Green New Deal or circular economy by integrating a natural capital perspective (e.g. by monitoring how actions implemented affect key environmental indicators). In this context, a single metric such as SESI shows the absolute performance of countries with regard to environmental sustainability and responds to the demands made from the 'Beyond GDP' community on the need for a single environmental sustainability metric that can complement GDP in its (mis-)use as a headline indicator for development.

In the future, feedback provided by different stakeholders as well as an increased availability of relevant data or scientific evidence that supports changes in existing environmental standards or the inclusion of different ones will require the structure and indicator selection of SESI to be revisited. Hopefully, the robustness of the framework and its potential applications will create the momentum for such review of the evidence and for relevant data to be generated.

5. Strong Environmental Sustainability Progress Index

5.1. Introduction

SESI – presented in chapter 4 – describes the performance of countries against science-based environmental standards at a given point in time. The index is intended to provide an intuitive message around environmental sustainability performance. The individual SES indicators, on the other hand, are meant to raise a flag in relation to the functions of natural capital that might be impaired, and therefore to show which issues demand policy attention. Nonetheless, SESI and the underlying indicators do not reflect whether the performance of countries is improving or worsening over time. To complement the snapshot view of SESI, I introduced SESPI in chapter 2, which is intended to provide a temporal perspective by showing whether countries are making progress towards environmental sustainability over time.

Historically, most environmental and sustainable development indices have – as SESI – reflected country performance at a given point in time (Hametner and Kostetckaia 2020). When time series were available for most indicators, progress was monitored by comparing the results of the latest year with those of previous years. Given that in most cases the metrics employed measured relative performance (i.e. the performance of countries against frontrunners), they failed to show systematically whether countries were making enough progress towards specific goals such as environmental policy targets or environmental standards. A few notable exceptions include the work of Sicherl (Sicherl 1973) and Ekins (Ekins and Simon 2001).

In the early 1970s, Sicherl (1973) proposed the time-distance approach as a way to complement the snapshot overview often presented by data users. The approach relies on two metrics: 'S-time-distance', which measures the time difference it takes a country to achieve a given level of a variable of interest reached by another country, and 'S-time-step', which shows the number of years needed in the past to increase one unit of a variable of interest (Sicherl 2011). In the context of the Sustainability Gap approach, Ekins and Simon (2001) proposed 'years to sustainability' (Y2S) in order to provide an easy-to-understand message about progress towards or away from environmental sustainability. Y2S represents the years required to reach a given environmental standard by linearly extrapolating current trends, thereby giving a general indication of whether countries are in the right track to achieve relevant environmental standards. Although easy to understand, the index presented a number of problems, the main one being the impossibility of aggregating the values when an indicator was showing negative trends and its Y2S value was infinite.

More recently, the emergence of the SDGs triggered new metrics intended to measure progress towards them, although most SDG indices still reflect country performance at a given point in time (see Hametner and Kostetckaia (2020) for some examples). In this context, Eurostat (2014a) provided an overview of methods that could be used to measure progress over time depending on the type of data available. This report laid the foundation for the annual series of 'Sustainable Development in the European Union' reports (Eurostat 2020b), whose methodology has been adapted in related research (Allen et al. 2020;

Hametner and Kostetckaia 2020; Simsek et al. 2020). Of especial interest in the context of the ESGAP framework is the method that compares observed trends with desired trends to evaluate not only whether countries are headed in the right direction, but also to evaluate whether, if maintained, observed trends would lead to reaching a given target at a given point in time. When a target value is available, Eurostat (2020b) and Sachs et al. (2020) use this method to measure progress towards the SDGs. Nonetheless, there are different ways of calculating observed and desired trends. In the case of Eurostat, observed and desired trends are assumed to follow an exponential function, while Sachs et al. (2020), on the other hand, use a linear function. The results of these assessments are presented in a variety of ways, most of which require normalising the data on trends to make it comparable across indicators. In the different publications, progress or lack thereof is commonly presented through the use a limited set of icons or colours to represent progress or lack thereof (Sachs et al. 2020), through a score-based system (Hametner and Kostetckaia 2020) or a combination of the two (Allen et al. 2020; Eurostat 2020b). Depending on the context, the comparison between observed and desired trends is interpreted at the level of individual indicators or at the level of indicator groups. The latter requires applying a normalisation, weighting and aggregation process to the results as with composite indicators.

Alternatives to this approach also exist in indicator-based assessments. The most notable one in Europe is the more qualitative perspective provided in the State and Outlook of the Environment Report (SOER) published by the European Environment Agency every five years (EEA 2019c). SOER provides the temporal perspective by combining data on trends, modelling results and expert input. Arguably, SOER-type assessments of trends are more comprehensive, but also demand a more complex process and require more resources to be implemented.

Against this background, this chapter computes SESPI for 28 European countries, thereby responding to the second research question: *Are European countries moving towards environmental sustainability?* The assessment presented here complements that in in chapter 4 by adding the previously missing temporal perspective. The construction of SESPI is also based on the manual on composite indicators (OECD and JRC 2008; JRC 2019) used previously. Nonetheless, since SESPI is an extension of SESI, the rationale behind some of the methodological choices remains the same and is therefore not elaborated on in the following sections. The most notable difference refers to the normalisation method, as instead of normalising country performance, SESPI normalises trend data. To that end, the Eurostat (2020b) methodology is used as starting point, as done by some of the latest assessments on SDG trends (Allen et al. 2020; Hametner and Kostetckaia 2020). In this context, the main novelty of SESPI is that instead measuring progress towards the SDG targets, it considers science-based environmental standards as goals to be reached. While the use of policy targets as reference adds, by definition, policy relevance to the assessment, policy targets are usually not aligned with science-based environmental standards (c.f. section 4.10). Hence, when policy targets are more lenient, progress metrics can provide a false sense of success, when in fact, progress towards environmental sustainability might be limited or even non-existent. For this reason, the use of SESPI provides more relevant insights on environmental sustainability from a strong sustainability perspective.

The chapter is structured as follows. Section 5.2 describes the methodology, while sections 4.9 and 5.4 present and discuss the results. Finally, section 5.5 concludes.

5.2. Methodology

5.2.1. Indicator selection

The indicators used to compute SESPI (hereinafter SESP indicators) are the same as the ones used to compute SESI. SES indicators have been selected based on their relevance, methodological soundness and data quality. By extension, this also holds true for the SESP indicators. As shown in chapter 4, for indicators to be relevant, they need to be related to the functions of natural capital, to have science-based environmental standards against which performance can be measured and to be meaningful at the national level. The methodological soundness criterion considers the readiness and maintenance of statistical production, accessibility and transparency, and compliance with existing methodological standards, while the data quality criterion covers aspects related to the frequency of dissemination, timeliness, time and geographical coverage and data comparability.

The list of indicators and data sources used to compute SESPI is shown in Table 30. SESPI retains the same general structure in terms of functions, principles and topics as the one shown in Figure 6. The main difference compared with Table 23 is the absence of the indicators on fish resources and access to green spaces because of lack of at least two broadly comparable data points that can be used to calculate trends. All in all, the index comprises 19 SESP indicators. Detailed information on each indicator can be found in Annex 1.

In this context, it is also important to bear in mind that there are some additional instances in which the temporal comparability of some datasets is limited. This is the case for the indicators measuring the chemical and ecological status of water bodies[12]. This is something already considered at the selection process, but should be kept in mind when interpreting the results.

[12] In its website, the European Environment Agency calls for caution when comparing countries' progress on the status of freshwater systems because the results are affected by the methods used to collect data (EEA 2018b).

Table 30: Final SESP indicator set

Function	Principle	Topic	SES indicator [Unit]	Data
Source	Renew renewable resources	Biomass	Forest utilization rate [%]	Forest Europe et al. (2015); Forest Europe (2020)
		Freshwater	Freshwater bodies not under water stress [%]	EEA (2018c)
			Groundwater bodies in good quantitative status [%]	EEA (2018b)
	Use non-renewables prudently	Soil	Area with tolerable soil erosion [%]	Panagos et al. (2015); Panagos et al. (2020)
Sink	Prevent global warming, ozone depletion	Earth system	CO_2 emissions [tonnes per capita]	Eurostat (2019a)
			ODS consumption [tonnes per capita]	Ozone Secretariat United Nations Environment Programme (2019)
	Respect critical levels and loads for ecosystems	Terrestrial ecosystems	Cropland and forest area exposed to safe ozone levels [%]	Horálek et al. (2019); Horálek et al. (2020)
			Ecosystems not exceeding the critical loads of eutrophication and acidification [%]	Tsyro et al. (2020)
		Freshwater ecosystems	Surface water bodies in good chemical status [%]	EEA (2018b)
			Groundwater bodies in good chemical status [%]	EEA (2018b)
		Marine ecosystems	Coastal water bodies in good chemical status [%]	EEA (2018b)
Life support	Maintain biodiversity and ecosystem health	Terrestrial ecosystems	Terrestrial habitats in favourable conservation status [%]	EEA (2020a)
		Freshwater ecosystems	Surface water bodies in good ecological status [%]	EEA (2018b)
		Marine ecosystems	Coastal water bodies in good ecological status [%]	EEA (2018b)
Human health and welfare	Respect standards for human health	Human health	Population exposed to safe levels of outdoor air pollutants [%]	Horálek et al. (2019); Horálek et al. (2020)
			Population using clean fuels and technologies for cooking [%]	WHO (2020)
			Samples that meet the drinking water criteria [%]	EC (2016)
	Conserve landscape and amenity	Other welfare	Recreational water bodies in excellent status [%]	EEA (2019f)
			Natural and mixed world heritage sites in good conservation outlook [%]	Osipova et al. (2017); Osipova et al. (2020)

5.2.2. Data treatment

5.2.2.1. Data gaps

Measuring progress over time requires at least two data points for each of the SESP indicators. These two data points need to be combined in a metric of temporal trends that can then be normalised, weighted and aggregated.

At this point, data-rich European countries face two problems in this respect. First, although the 19 indicators that form SESPI have more than one data point, the years for which data is available vary widely. For instance, an indicator might have data for the period 2011-2013 while another could have data for the years 2005 and 2015. While this does not prevent the calculation of a progress metric, it is an important caveat that should be stated clearly. This problem will likely remain in the near future, given that the frequency with which the data of the indicators used here is compiled depends on several factors such as the producer (e.g. statistical offices, researchers, environmental agencies) or the legislation in place. Second, beyond the different temporal data availability, some indicators have data gaps as shown in Table 24 in the previous chapter. Before calculating temporal trends for each SESP indicator, these two problems need to be addressed.

Regarding the first problem, short-term and long-term trends should be calculated to provide a more complete picture of progress over time. As argued in the previous paragraph, this is not possible because the years for which data is available differs between indicators. In this first version of SESPI, only short-term periods have been considered for conveniency issues. When possible, short-term is defined as a five-year period since the last available data point, but changes might sometimes be required depending on data availability. The years considered for each indicator are shown in Table 31.

Table 31: Years used to compute trend in SESP indicators

SESP indicator	Year 0 (t_0)	Year 1 (t_1)
Forest utilization rate	2010	2015
Freshwater bodies not under water stress	2010	2015
Groundwater bodies in good quantitative status	2009	2015
Area with tolerable soil erosion	2010	2016
CO_2 emissions	2013	2018
ODS consumption	2014	2019
Cropland and forest area exposed to safe ozone levels	2012	2017
Ecosystems not exceeding the critical loads of eutrophication and acidification	2005	2017
Surface water bodies in good chemical status	2009	2015
Groundwater bodies in good chemical status	2009	2015
Coastal water bodies in good chemical status	2009	2015
Terrestrial habitats in favourable conservation status	2012	2018
Surface water bodies in good ecological status	2009	2015
Coastal water bodies in good ecological status	2009	2015
Population exposed to safe levels of outdoor air pollutants	2012	2017
Population using clean fuels and technologies for cooking	2013	2018
Samples that meet the drinking water criteria	2011	2013
Recreational water bodies in excellent status	2014	2019
Natural and mixed world heritage sites in good conservation outlook	2017	2020

As for data gaps, when computing SESI it was deemed appropriate to fill the data gaps of the most recent year to get an estimation of an indicator value. Filling data gaps for two different years and calculating the resulting trend introduces much more uncertainty and therefore has not been deemed appropriate in the context of SESPI. Thus, instead of estimating the data gaps for the first year in the table above, the indicators for which no trend could be calculated because of a missing data point have been excluded from the sample. All available data points were considered valid, and therefore there was no need to correct outliers that often distort the sample distribution.

5.2.2.2. Observed and desired trends

Eurostat (2014a) describes different methods to measure progress. In the context of the SDGs, Eurostat (2020b) uses two of them. When a quantitative target is available, it compares observed trends with desired trends, the latter representing the theoretical trend that would lead to achieving the SDG target in 2030. When a quantitative target is not available, they use arbitrary threshold values to classify trends in different groups.

SESPI uses the first method, since all the SESP indicators have a science-based environmental standard. To that end, the data of the years shown in the previous table are combined to calculate the linear trends of each indicator (as opposed to the exponential trends as in Eurostat). The formulation of the annual change ($trend_{obs}$) for a period going from t_0 (base year) to t_1 (most recent year) is given below:

Equation 5

$$trend_{obs} = \frac{I_{t_1} - I_{t_0}}{t_1 - t_0}$$

where I represents the value of each indicator at a given point in time.

Because on its own, annual changes are not enough to assess whether enough progress towards environmental sustainability is being made, observed trends are compared to the desired trends ($trends_{des}$), which represents the change needed to reach a target (in this case an environmental standard) in a given year. Desired trends are calculated as follows:

Equation 6

$$trend_{des} = \frac{x_{t_r} - I_{t_1}}{t_r - t_1}$$

where x is the target value (100 in most indicators) and t_r is the target year. The choice of the latter is arbitrary; it could be ten, 20, 30 years or a specific year that has a political meaning. In this exercise, t_r is set to 2030, given its relevance in the context of the SDGs.

The ratio between observed and desired trends (R_{o-d}) provides an intuitive metric of whether enough progress is being made in each individual indicator.

Equation 7

$$R_{o-d} = \frac{trend_{obs}}{trend_{des}}$$

Negative scores for R_{o-d} indicate that country performance is worsening and therefore it will be impossible to reach the environmental standard unless those trends are reversed. Values higher than 100% suggest that under current trends the environmental standard will be met before the target year, while values between 0% and 100% are indicative of an improving trend that is still insufficient to meet the environmental standard by the target year.

5.2.3. Normalisation

Normalised country scores depend on the difference between the observed and desired trajectory. Thus, indicators in which observed trends are close to those considered sustainable will get high normalised scores, while indicators in which observed trends are not aligned with desired trends will get low normalised scores. In order to formalise the mathematical formulation of the statement above, we use the goalpost normalisation method. In the goalpost method, the user defines upper and lower goalposts aligned with sustainable and unsustainable conditions, which are then assigned a normalised score of 100 and 0 respectively. In practice, there are two slightly different approaches depending on the type of indicator.

In indicators that represent an environmental or social state bound in the 0-100% range, the normalisation is carried out as shown in Table 32. Generally speaking, the normalisation can be interpreted as follows:

- When an environmental standard has not been met in t_1, a normalised score of 100 reflects that, if continued, current trends would lead to meeting the environmental standard in 2030 (the reference year). A score of 50 is assigned when no (positive or negative) progress occurred between t_0 and t_1. Positive trends that are insufficient to meet the environmental standard are scored between 50 and 100. On the negative side, a negative trend that mirrors the positive trend needed to reach the environmental standard is assigned a score of zero. In between, scores between zero and 50 are assigned to less negative trends.
- When an environmental standard has been reached in t_1, $R_{o\text{-}d}$ is the result of dividing by zero and therefore is problematic. With maintenance of the standard, or further improvement, the environmental standard would also be met in 2030. This is reflected through a normalised score of 100. Given that the maximum value of the indicator and the environmental standard are equal (i.e. 100), the environmental standard cannot be reached in t_1, under worsening trends.

Table 32: Normalisation of environmental and social state SESP indicators

Situation	**Trend needed**	**Actual trend**	**Normalisation**
Environmental standard met in t_1	$trend_{des} = 0$	$trend_{obs}$ is always ≥ 0. Under current trends, the environmental standard will be met in 2030.	$NI = 100$
Environmental standard not met in t_1	$trend_{des} > 0$	Depending on the evolution, $trend_{obs}$ can have positive, zero or negative values. Meeting the environmental standard in 2030 depends on the $R_{o\text{-}d}$ value.	$if \quad R_{o-d} \geq 1 \quad NI = 100$ $if \quad -1 < R_{o-d} < 1 \quad NI = 50 + 50R_{o-d}$ $if \quad R_{o-d} \leq -1 \quad NI = 0$
	$trend_{des} < 0$ It does not apply	It does not apply	It does not apply

In indicators that represent environmental pressures or that are not bound in the 0-100 range (e.g. CO_2 emissions, consumption of ODS and forest utilisation rate), the normalisation differs slightly as shown in Table 33, although the logic remains largely the same. In these indicators, environmental sustainability increases when CO_2 emissions, consumption of ODS and forest utilisation rates decrease. Two cases arise: when environmental standards have not been met in t_1 and when they have been met.

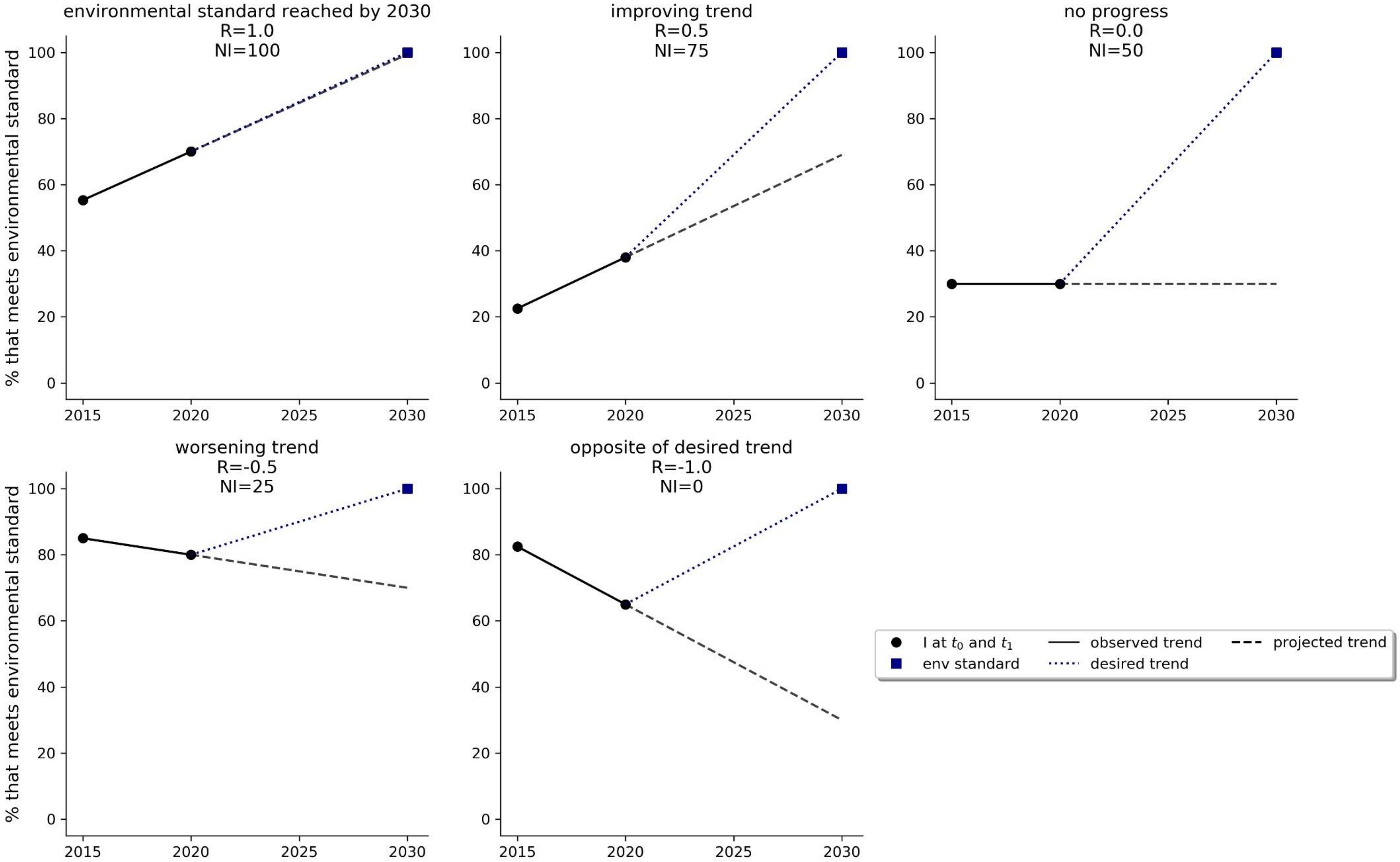
- When an environmental standard has not been met in t_1, which covers most of the cases, the exact same logic as in Table 32 applies. The only difference is that in these indicators, both observed and desired trends will have a negative sign, since a decrease in the indicator value represents an improvement.
- In some cases, the indicator value is lower than the environmental standard, which indicates that the country performs better than what is required. (e.g. CO_2 emissions or forest utilisation rates can be lower than the environmental standard in t_1). Thus, an improving trend or the absence of change will always lead to meeting the environmental standard in 2030. In a situation in which trends worsen, meeting the environmental standard will depend on the value of $R_{o\text{-}d}$ as shown in the table.

Table 33: Normalisation of remaining SESP indicators

Situation	Trend needed	Actual trend	Normalisation
Environmental standard met in t_1	$trend_{des} > 0$	Depending on the evolution, $trend_{obs}$ can have positive, zero or negative values. Meeting the environmental standard in 2030 depends on the R_{o-d} value.	$if \quad R_{o-d} \leq 1 \quad NI = 100$ $if \; 1 < R_{o-d} < 2 \quad NI = 50 + 50(2 - R_{o-d})$ $if \quad R_{o-d} \geq 2 \quad NI = 0$
Environmental standard not met in t_1	$trend_{des} < 0$	Depending on the evolution, $trend_{obs}$ can have positive, zero or negative values. Meeting the environmental standard in 2030 depends on the R_{o-d} value.	$if \quad R_{o-d} \geq 1 \quad NI = 100$ $if \; -1 < R_{o-d} < 1 \quad NI = 50 + 50R_{o-d}$ $if \quad R_{o-d} \leq -1 \quad NI = 0$

The normalisation process is visually represented in Figure 23, where the values of a fictional indicator are shown for five fictional countries (see the note at the bottom of the figure). A worked example for the 19 SESP indicators for Europe as a block is presented in Table 42 in Annex 2.

Figure 23: Interpretation of the normalised scores for a fictional SESP indicator in different fictional countries



In the first country (top left), observed and desired trends are equal and therefore, the environmental standard will be reached in 2030 under current trends ($R_{o\text{-}d}$=1), which gives a normalised score of 100. In the second country (top centre), the observed trend shows a change in the right direction ($R_{o\text{-}d}$=0.5), but this will be insufficient to meet the environmental standard by 2030. In the third country (top right), there is no progress ($R_{o\text{-}d}$=0), which leads to a normalised score of 50. In the fourth country (bottom left), change occurs in the wrong direction ($R_{o\text{-}d}$=-0.5), which leads to a normalised score of 25. Finally, in the last country (bottom centre), observed change is the opposite of what it should be to meet the environmental standard ($R_{o\text{-}d}$=-1). This is equivalent to normalised score of zero.

5.2.4. Weighting and aggregation

In order to align the meaning of SESI and SESPI, the construction of the latter needs to be consistent with that of the former. Thus, equal weights and a weighted geometric mean are used in the weighting and aggregation processes. As with SESI, zeros and small values are treated to avoid the problems arising from their presence when aggregating with the geometric mean. The rationale behind their use is extensively described in the previous chapter.

The resulting progress index can be interpreted in a similar vein as SESI. A value of 100 indicates that all the indicators describe trends that are aligned with meeting their respective environmental standards in 2030. A score of zero, indicates that all the indicators are going in the wrong direction and, therefore, in 2030 the environmental sustainability performance of countries will have deteriorated considerably. In between, low scores suggest that a (at least) a few indicators are going in the wrong direction, and therefore several environmental functions will be threatened in the future. High scores reflect the opposite.

As with SESI, zeros and small values are treated to avoid the problems arising from their presence when aggregating with the geometric mean. Thus, a minimum score of five is assigned to all the normalised values before aggregation.

5.2.5. Statistical and conceptual coherence

The conceptual coherence analysis seeks to understand how the choices made during the construction of an index are aligned with its theoretical framework. In this context, it is important to bear in mind that SESPI is intended to mirror SESI to the extent possible. For this reason, SESPI builds on, whenever possible, the same indicators as SESI and follows the same logic in the normalisation, weighting and aggregation processes. Thus, SESPI will be as aligned with the ESGAP framework as SESI is.

The results of the conceptual coherence analysis of SESI reported in Annex 1 showed that the interpretation of the relevance criteria in the selection of indicators and the use of the goalpost normalisation method are consistent with the definition of environmental sustainability, which requires the use of environmental standards to represent the conditions under which the functioning of natural capital can be maintained over the long term. The choice of the aggregation method and the approach selected to treat zeros and small values is linked to substitutability of the functions of natural capital. Last, the selection of equal weights deviates from the notion that life support functions are more relevant than the other sets of functions, but the lack of an adequate weighting method prevented this aspect from being addressed properly.

The goal of the statistical coherence analysis is to understand how the information is translated from the indicators to the final index with the ultimate goal of revising the structure and key choices made in the construction. Because the construction of SESPI is largely defined by SESI, a statistical coherence analysis will offer limited insights and has therefore been discarded.

5.2.6. Uncertainty analysis

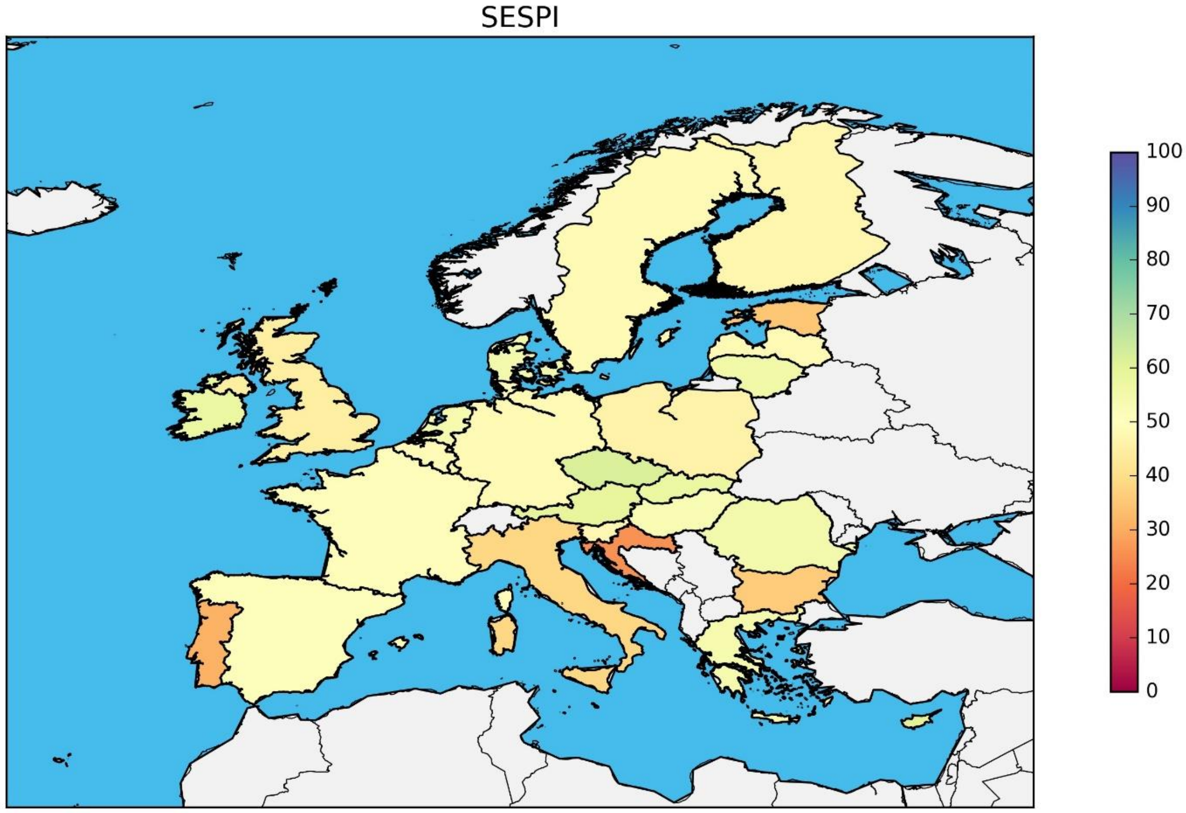
The uncertainty analysis of SESI has tested how sensitive the scores of the different dimensions of SESI and the normalised scores of the indicators are to choices made in the normalisation, weighting and aggregation processes. The uncertainty analysis, the results of which are reported in section 4.9.5, showed that indeed, SESI is very sensitive to choices in the construction of the index, and therefore those choices need to be aligned with the theoretical underpinnings of the ESGAP framework, as it is currently the case.

Given that testing the same assumptions would not provide any new insights, uncertainty to a more relevant variable is tested here: time. As shown in Equation 7, the starting point to calculate the normalised scores are linear trends. Although there is a rationale behind the choice of the year used as t_0 in the denominator – the baseline year –, ultimately other time points could be selected if data were available. To understand how sensitive the SESPI results are to this assumption, the baseline year has been selected randomly from all the years for which data was available. Thus, the index and function scores have been calculated for 1,000 different combinations selected through a Montecarlo analysis. To understand the effects of the baseline year in individual SESP indicators, the normalised scores have also been computed with all the potential combinations of t_0, while keeping t_1 – the last year for which data was available – constant.

5.3. Results

Figure 24 shows the SESPI scores of the 28 European countries covered in this paper. Most countries score between 40 and 60 points, which suggests that under current trends they will not reach all the environmental standards in 2030, the closing year of the SDGs. In the top, the Czech Republic, Luxembourg and Latvia are slightly above the 60-point line. This can be interpreted as most indicators moving in the right direction, with only a few showing no progress or going in the wrong direction (it should be remembered that using the geometric mean of the indicators for aggregation gives greater weight to the lower indicator scores, to reflect the non-substitutability characteristic of strong sustainability). In this context, it is important to bear in mind that, for individual SESP indicators, a normalised score of 100 indicates that under current trends an indicator will achieve the environmental standard in 2030 or sooner, or it has already achieved it. A score of 50 shows that no progress has been reported, while a score of zero shows that current trends are exactly the opposite of what is needed to meet the environmental standard in 2030. At the bottom, Italy and Portugal have less than 34 points, and Croatia gets a score of 26. These countries will not only miss the environmental standards but are also going in the wrong direction in many instances. The European block scores 42 points. The reader should note that in exceptional cases, data gaps result in the scores of some countries being computed with slightly fewer indicators.

Figure 24: SESPI score for European countries



SESPI scores are easier to interpreted when shown together with SESI scores (Figure 25). The goal in the figure is to be in the blue box in the upper right corner. Nevertheless, this is not the case for any of the countries. Ireland is the only country in the yellow shade of the figure with SESI and SESPI scores of around 60. Most countries score under 50 in both indices, with a few scoring above 50 only in one of the two. Broadly speaking, it can be argued that European countries are far from being environmentally sustainable or making enough progress to be environmentally sustainable in 2030. Of course, different indicators show different trends and, therefore, SESPI scores need to be complemented with the scores obtained at lower levels. There is no apparent correlation between SESI and SESPI scores (R=-0.04, p-value<0.01).

Figure 25: SESI and SESPI scores for European countries

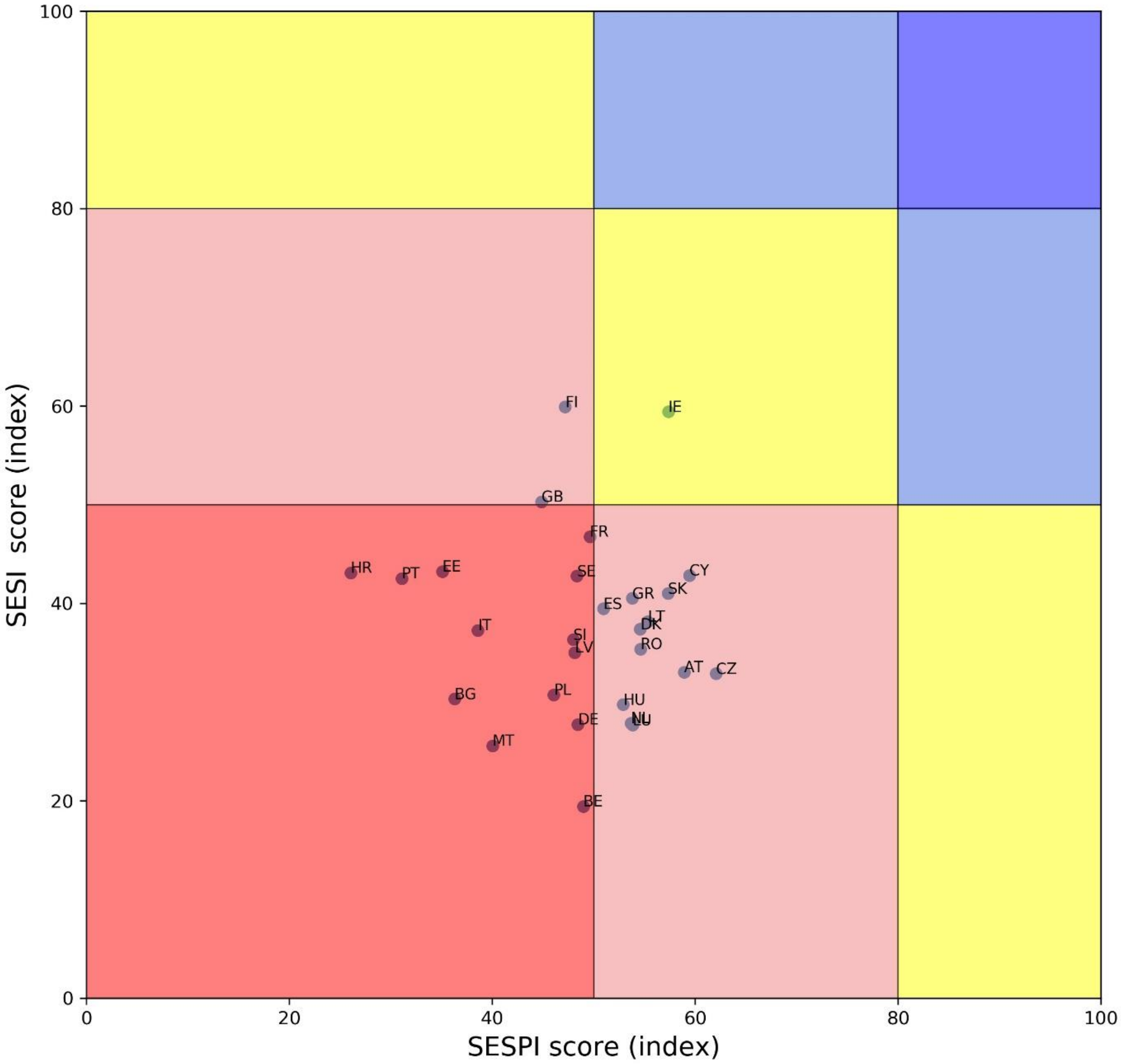


Figure 26 and Figure 27 represent a heatmap of the SESPI scores at the level of environmental function and sustainability principle. Countries perform worse in the source function, which considers the provision of biotic and abiotic resources. In this function, scores range from 71 in Lithuania to 16 in Portugal, with 22 countries scoring less than 50 points. The European block scores 24 points. The overall score of the source functions is driven down mainly by the low performance of two indicators of renewable resources: forest resources and freshwater resources. In the former, although many Northern and Central-West European countries experienced an increase in the net annual increment of forest resources between 2010 and 2015, fellings increased at a higher rate, which led to higher exploitation rates and therefore a worsening trend (Forest Europe 2020). In South-East Europe, available resources barely changed in the same period, but fellings increased, thereby resulting in higher exploitation rates as well. In Central-Eastern European countries exploitation rates decreased. With regard to freshwater resources, the river basin areas suffering from water stress in at least one quarter of the year increased between 2010 and 2015. This is partly the result of lower available freshwater resources in 2015 due to a significant decrease in net precipitation (Eurostat 2021). Performance in groundwater scarcity is generally much better. Between 2009 and 2015, the area of

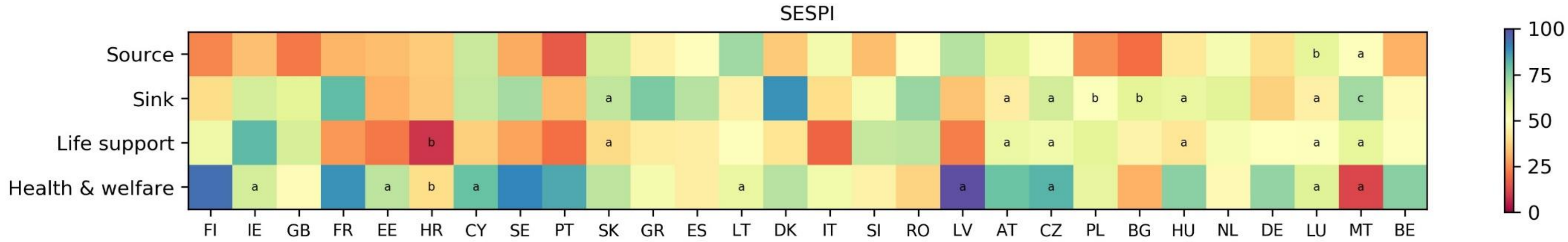
European groundwater bodies in good quantitative status increased from 87% to 90%, which results in a normalised score of 100. This follows a continued decrease in groundwater abstraction in Europe since 1990 (EEA 2019c). At the country level, trends are generally good with more than half of the countries headed towards achieving the environmental standards by 2030. In the case of soil erosion, at the European level there has been barely any change in the area that is subject to tolerable soil erosion rates. This is partly because erosion rates in arable lands tend to be much higher than the environmental standard, and therefore, even when erosion rates are reduced, the percentage of land area that meets the environmental standard might not increase. Nonetheless, Panagos et al. (2020) report positive signs as a result of conservation practices in countries such as Austria, Denmark, Germany, Estonia, France and Portugal. On the other end, they mention Bulgaria as a laggard in the implementation of management practices intended to reduce soil erosion. Perhaps most worrying, the performance of some Mediterranean countries that suffer from high erosion rates has worsened between 2010 and 2016.

The European block reports scores between 48 and 55 in the remaining functions with relevant differences in the underlying principles. In the sink functions, scores tend to be higher for Earth System processes with all the countries scoring 100 in ODS consumption (where the standard has already been met) and many countries reporting progress in reducing CO_2 emissions. In this vein, although 18 European countries reported average annual per-capita CO_2 emission reductions in the range of 0-11% between 2013 and 2018, these are in most cases not sufficient to meet the environmental standard in 2030. As a result, most normalised scores range between 50 and 90. The remaining 10 countries reported increases in emissions between 0-3%. Regarding chemical pollution in ecosystems, country performance is much more uneven with France, Denmark and Romania generally moving in the right direction, and 15 countries obtaining scores below 40. The European block shows improving trends in the chemical status of terrestrial ecosystems (stronger in relation to ozone pollution compared to eutrophication and acidification). In contrast, small progress was reported in the chemical status of groundwater, while the situation of surface and coastal water systems worsened. The reader should note that the latter statement needs qualifications on two grounds. First, although the percentage of surface and coastal water bodies in good chemical status decreased between 2009 and 2015, significant progress has been made in reducing the concentration of some pollutants such as pesticides or some heavy metals (EEA 2018b). Nonetheless, the presence of other substances such as mercury leads to failure to meet good chemical status in numerous freshwater bodies (EEA 2018b). Second, caution is advised when comparing the country performance over time, as the results are affected by the methods used to collect data, which might differ.

In the life support functions, Ireland, Romania and Slovakia are at the top, while 14 countries score less than 50 points. At the European level, progress is similar across the three broad ecosystem categories considered (terrestrial, freshwater and coastal) with scores that range between 44 and 52. In terrestrial ecosystems, the percentage of habitats classified as having a good conservation status decreased slightly between 2012 and 2018. Trends differ considerably depending on the country and terrestrial habitat type (EEA 2020b). Freshwater and coastal ecosystems describe a relatively stable situation with a very small change between 2009 and 2015 at European level, with high variation between countries (EEA 2018b). As in the previous paragraph, the trends reported should be interpreted carefully because of the methods used to assess the ecological status of freshwater ecosystems. Beyond comparability issues, it seems clear that under these trends, terrestrial, freshwater and coastal ecosystems will not meet the environmental standards by 2030. This is specially worrying in the case of terrestrial and freshwater ecosystems, where only 16% and 36% of the ecosystems met the standard in the last year for which data was available.

Lastly, most European countries report progress in the human health and welfare functions with 14 countries scoring more than 75 points, three of which with a normalised score of 100. The European block scores 52 points. The country distribution of the scores in indicators of human health, on the one hand, and other welfare aspects, on the other, is similar, although countries with high scores in one of the principles do not necessarily have high scores in the other. When it comes to indicators related to human health, the European block shows mixed progress. While the percentage of population exposed to outdoor air pollution levels below the WHO guideline values more than doubled from 11 to 26 between 2012 and 2017 (score 76), the population with access to clean cooking fuels declined slightly (score 14), although most of the population meets the environmental standard. In the drinking water indicator, the European block obtained a score of 100. With regard to other welfare functions, the number of European bathing sites reporting excellent water quality increased from 86% to 88% between 2014 and 2019. At this pace, the environmental standard would not be reached by 2030. At the national scale, ten countries reported progress compatible with meeting the environmental standard in the near future, while nine others reported some progress, although insufficient. Last, there have barely been any changes in the conservation status of natural and mixed World Heritage sites between 2017 and 2020. Accordingly, most countries obtain a score of 50, while the European block scores 52 points.

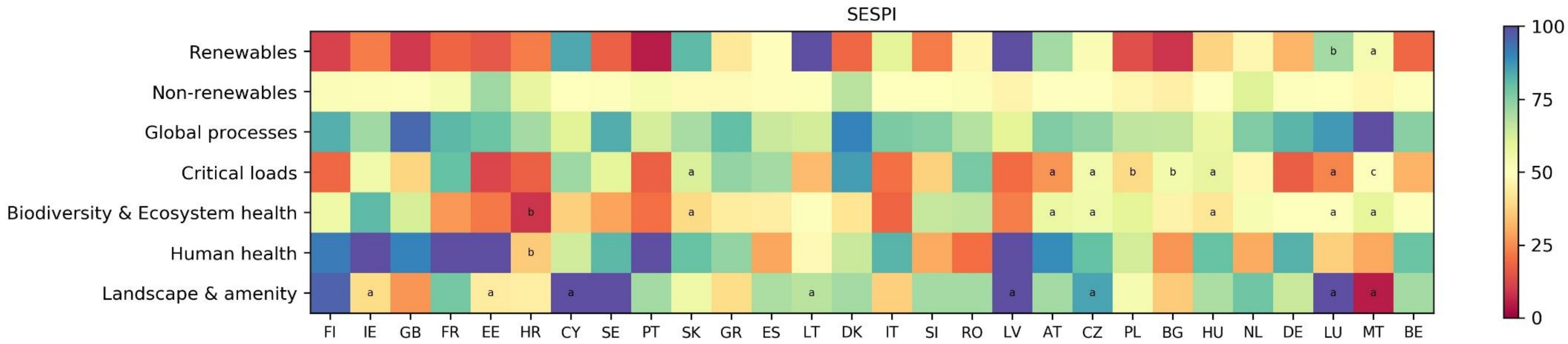
Figure 26: SESPI scores by environmental function



The figure shows the scores of each country for the four environmental functions. Dark red indicates low scores, while dark blue indicates high scores. Countries are sorted by the SESI score from higher to lower.

The label [a] in the heatmap indicates that one of the indicators assigned to the function is blank because it does not apply to the country (e.g. coastal areas in landlocked countries). The labels [b] and [c] indicate that one and two indicators do not have enough data to be integrated in SESPI. These gaps come on top of those for the indicators on fish resources and access to green areas.

Figure 27: SESPI scores by sustainability principle

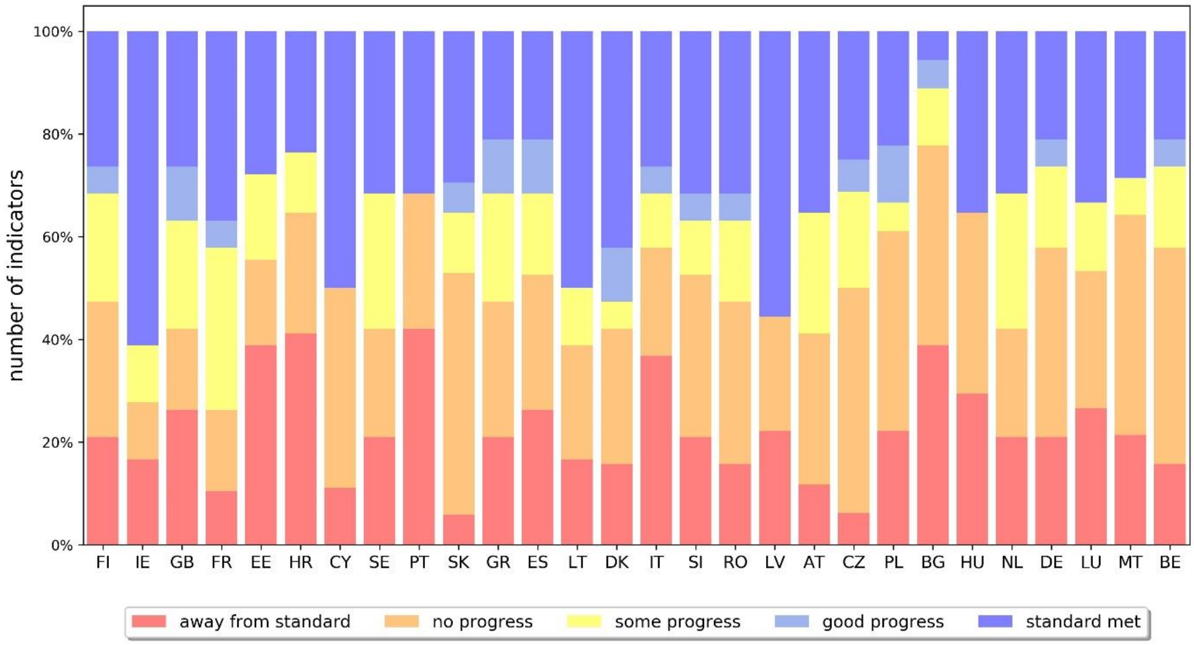


The figure shows the scores of each country for the seven sustainability principles. Dark red indicates low scores, while dark blue indicates high scores. Countries are sorted by the SESI score from higher to lower.

The label [a] in the heatmap indicates that one of the indicators assigned to the principle is blank because it does not apply to the country (e.g. coastal areas in landlocked countries). The labels [b] and [c] indicate that one and two indicators do not have enough data to be integrated in SESPI. These gaps come on top of those for the indicators on fish resources and access to green areas.

Figure 28 shows country performance by the progress reported in individual SESP indicators. In Europe (not shown in the figure), 32% of the indicators are moving away from the environmental standard, 32% show no progress, 21% describe some progress and 16% will meet the environmental standard if observed trends are maintained. As expected, there is a strong negative correlation between the index score and the percentage of indicators that are moving away from the environmental standard ($R=-0.78$, $p<0.01$). This is the result of the geometric mean driving down the index score by penalising poor performances.

Figure 28: Progress reported by SESP indicator



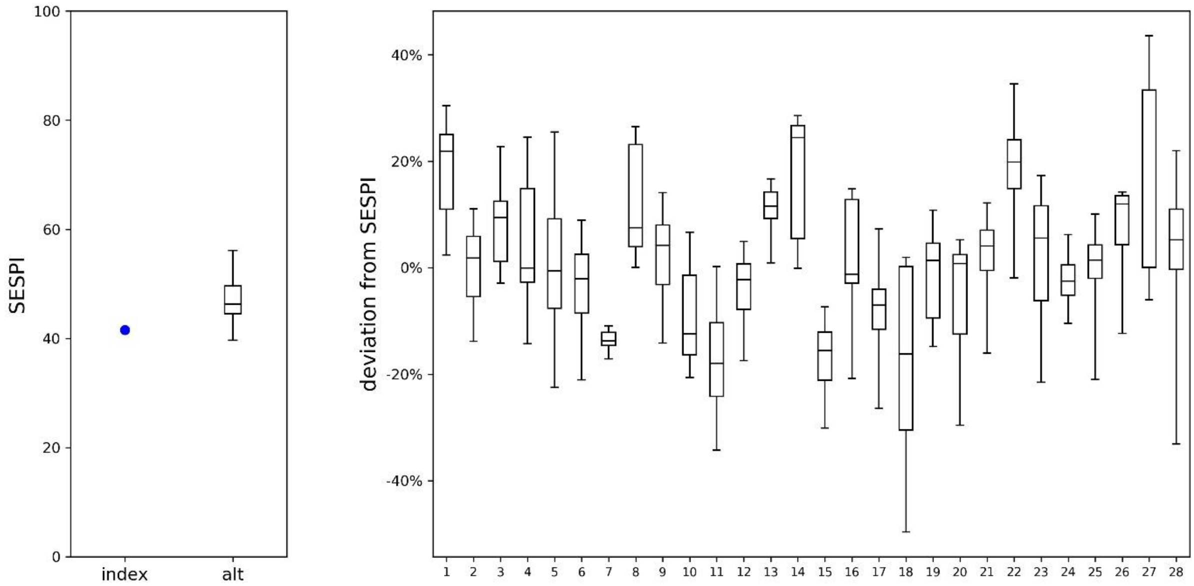
The figure shows the progress made by each country in meeting the environmental standards of all the indicators for which data is available. Dark red indicates scores below 45, orange scores between 45 and 55, yellow scores between 55 and 79, light blue scores between 80 and 95, and dark blue scores between 95 and 100. Countries are sorted by the SESI score from higher to lower.

5.3.1. Uncertainty analysis

As in the case of SESI, it is important to understand how the assumptions made during the construction of the index affect the results. The effects of the normalisation, weighting, aggregation and treatment of zeros and small values have been assessed in section 4.9.5. Since SESPI is based on the temporal trends of indicators, uncertainty to the choice of the specific time points used to calculate observed and desired trends is presented below.

When using the years shown in Table 31 to calculate trends, the score for the European block is 42. Using different data points as t_0 generally leads to higher index scores (median 46) as shown in the left side of Figure 29: Uncertainty associated with time in SESPI at index level. At the country level, in most cases changes in index scores range from ±20%, although exceptions apply.

Figure 29: Uncertainty associated with time in SESPI at index level



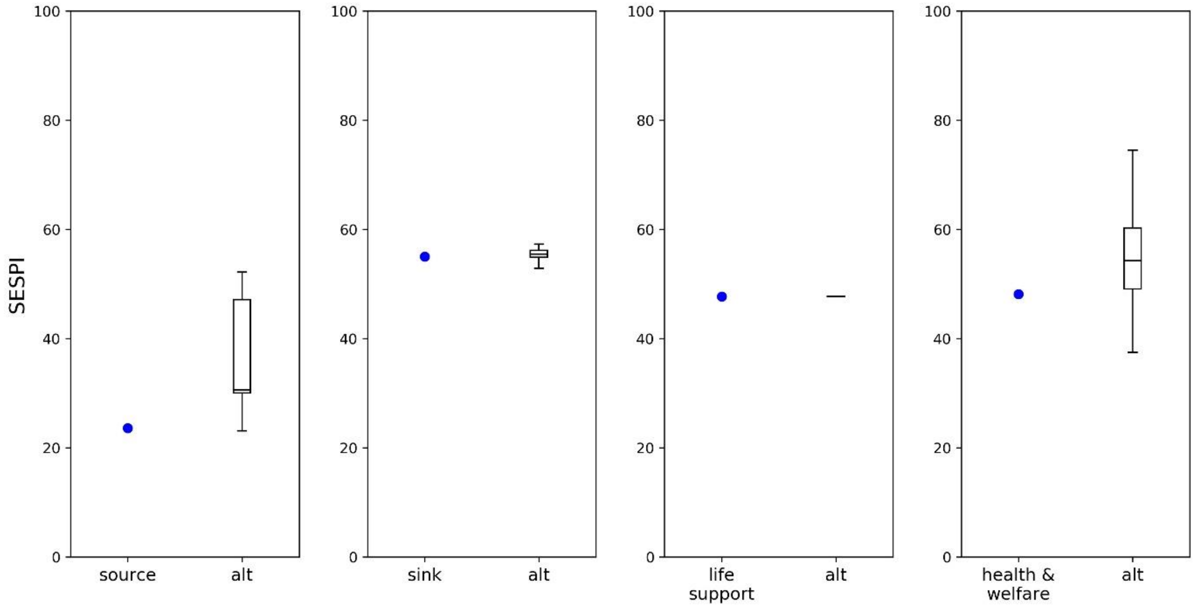
The figure on the left shows the index score at European level with the default and alternative base years. The figure on the right compares the ranks obtained with the default and alternative base years.

The x axis in the first figure shows the default and the alternative values generated using different data points as t_0. alt: alternative. The x axis in the second figure represents the 28 European countries ordered by SESI score.

The upper and lower edges of the rectangle in the boxplot represent the 75th and 25th percentiles, while the top and bottom markers represent the maximum and minimum values.

The differences by broad function category differ considerably for the European block as shown in Figure 30, which compares the default scores at European level with those obtained using different time points. Source and human health and welfare are the most affected functions. The score for the source functions tends to be higher (24 with the default method, median of 31 with alternative) with virtually all the runs leading to a higher score. In the case of human health and welfare functions, the median score obtained in the Montecarlo analysis is similar to the default score (54 and 52 respectively), although much higher and lower scores are obtained depending on the run. The default and alternative methods in the sink and life support functions yield very similar results. In the case of life support functions, the same score is obtained in every run. The reason is that the indicators in this category only have two data points, so no real alternative could be tested. Something similar occurs in the sink functions, where four out of seven indicators only have two data points. The rest show relatively constant changes irrespective of the time point used as t_0.

Figure 30: Uncertainty associated with time in SESPI at function level

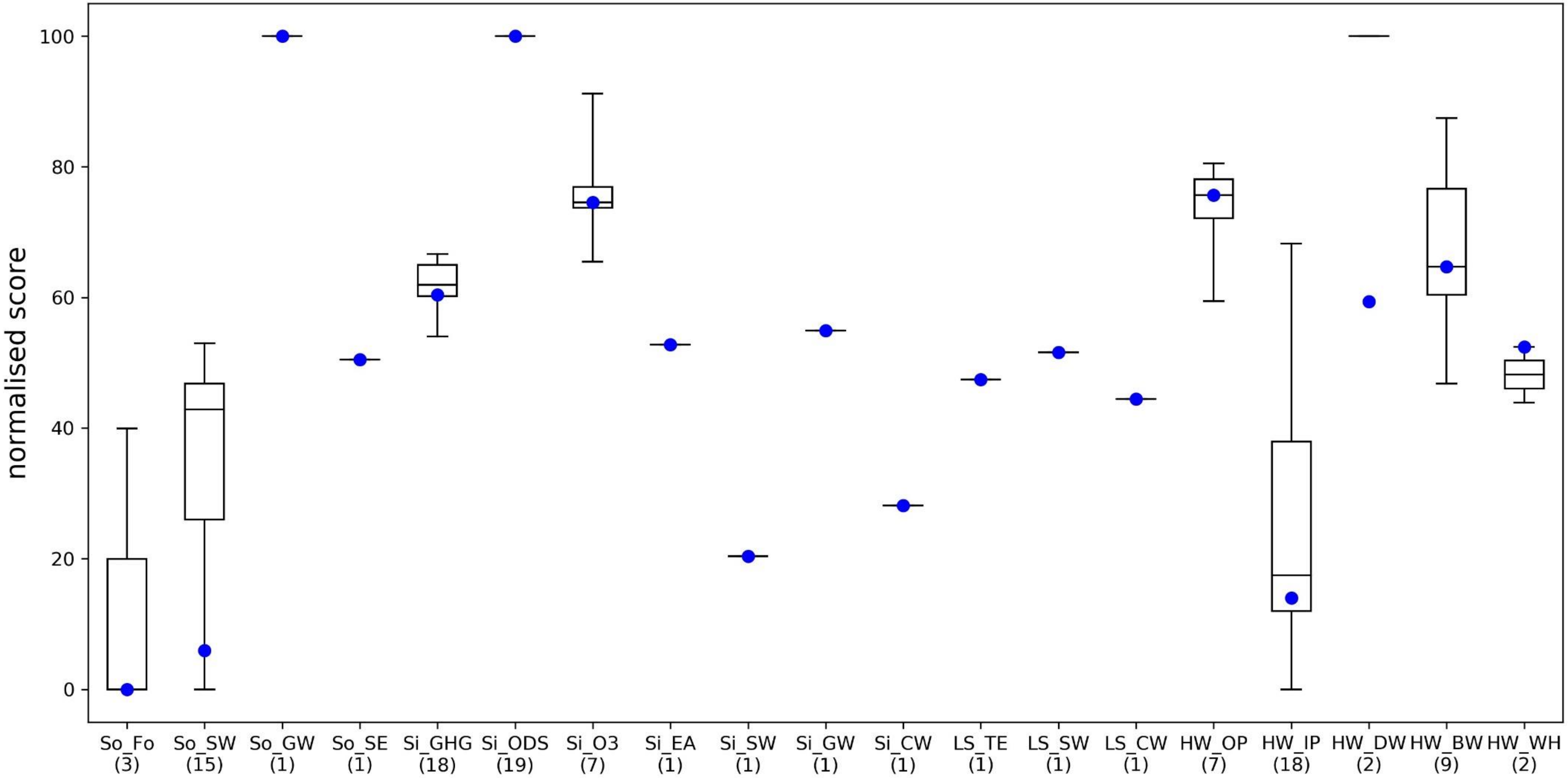


The x axis in the figures shows the default and the alternative values generated using different data points as t_0. alt: alternative. The scores are shown for Europe as a block.

The upper and lower edges of the rectangle in the boxplot represent the 75th and 25th percentiles, while the top and bottom markers represent the maximum and minimum values.

The latter more evident in Figure 31, which compares the normalised scores of each indicator to the normalised scores obtained using all possible data points as baseline years. Nine of the 19 indicators only have two data points, which gives only one combination to calculate observed and desired trends, and therefore results in the same score as the default option. In the source functions, two indicators have only two data points, while the other two seem quite sensitive to the choice of the baseline year. Thus, while the normalised score of the forest resources and freshwater scarcity indicators changes considerably for the European block, in both cases the progress reported is null or negative (i.e. it has a normalised score of 50 or less). As already noted, four out of seven indicators in the sink functions have only two data points. The other three report relatively similar progress irrespective of the baseline year used. In the life support functions all three indicators have only two data points. Lastly, the human health and welfare indicators show the largest sensitivity to the baseline year chosen. In this context, it is important to note that, except for the indoor air quality indicator, the other indicators are consistent in reporting progress or lack thereof, although the intensity of the progress varies considerably depending on the baseline year chosen.

Figure 31: Uncertainty associated with the selection of t0 in SESPI at indicator level



The blue dots show the normalised score obtained with the t_0 and t_1 values in Table 31. The upper and lower edges of the rectangle in the boxplot represent the 75th and 25th percentiles, while the top and bottom markers represent the maximum and minimum values.

The values in parenthesis under each indicator acronym represent the number of t_0-t_1 combinations available to compute observed and desired trends.

So_Fo: Forest utilization; Surface water scarcity, So_GW: Groundwater scarcity; So_SE: Soil erosion; Si_GHG: CO_2 emissions; Si_ODS: Consumption of ODS; Si_O3: Ozone pollution in terrestrial ecosystems; Si_EA: Eutrophication and acidification in terrestrial ecosystems; Si_SW: Chemical pollution in surface waters; Si_GW: Chemical pollution in groundwater; Si_CW: Chemical pollution in coastal waters; LS_TE: Ecological health of terrestrial ecosystems; LS_SW: : Ecological health of surface waters; LS_CW:

Ecological health in coastal waters; HW_OP: Outdoor air pollution; HW_IP: Indoor air pollution; HW_DW: Drinking water quality; HW_BW: Quality of bathing waters; HW_WH: Conservation of World Heritage sites.

5.4. Discussion

5.4.1. Measuring progress towards environmental sustainability

Environmental and sustainable development metrics have historically provided a snapshot perspective, thereby informing about country performance at a given point in time. Although metrics intended to capture temporal trends have been around for a long time (e.g. Sicherl (1973); Ekins and Simon (2001)), recently this dimension has gained more importance through the SDG-related metrics (Eurostat 2020b; Sachs et al. 2020).

Beyond assessing whether the functions of natural capital are threatened, the need to provide insights on whether countries are moving in the right direction has been a key aspect of the ESGAP framework since its inception (Ekins et al. 2003b). In order to address this aspect and to complement the snapshot perspective given by SESI, SESPI intends to shed light on whether countries are making enough progress towards or away from environmental sustainability. To that end, SESPI shares the same structure as SESI and mirrors, to the extent possible, its set of indicators, but instead of reflecting whether environmental standards are met in a given year, the data is used to compare observed trends with those required to meet the environmental standards sometime in the future (in this case 2030). The data produced for this comparison is then normalised and aggregated, following the weighting of the indicators, into a single score, where an index value of 100 indicates that, if sustained, the trends reported for each indicator would lead to meeting all the environmental standards by 2030. Conversely, a score of zero indicates that for every indicator the change needed to achieve environmentally sustainability is occurring in the wrong direction. In between, high scores represent improving trends for most indicators, while low scores indicate the opposite. While interpreting the results, it is important to bear in mind that the index cannot be considered a forecast of the future, since it does not indicate whether those trends will actually be sustained.

SESPI is intended to complement SESI. A statistical analysis suggest that this is actually the case, since there is no correlation or limited correlation between the SESI and SESPI scores at index and function levels (R=-0.04 for index, R=-0.14, R=0.17, R=-0.07 and R=0.45 for source, sink, life support and human health and welfare functions respectively). When combined, both indices can be used to create appealing narratives around the environmental sustainability performance and trends of countries. Arguably, SESPI is less intuitive than SESI because of the meaning of the normalised scores. In this vein, a score of 100 always indicates that an indicator is on track to reach its environmental standard by 2030. Nonetheless, since desired trends differ between indicators (with some even requiring no change if the environmental standard is met in the present), different growth and decline rates will lead to different normalised scores. Thus, as a general rule, normalised scores are defined by the context, rather than by the absolute value of the change reported by countries, similar to the approach used by Eurostat (2020b). In order to better describe the information on trends, additional information on the number of indicators describing positive, negative or no change can be used alongside the index and indicator scores.

5.4.2. Are European countries moving towards environmental sustainability?

European countries show mixed progress towards environmental sustainability. Europe as a block scores 42 points with relevant differences between environmental functions and indicators. The highest score in an environmental function is 55, far from the scores that would indicate substantial progress towards meeting the environmental standards in the near future.

Europe is making little progress in the management of natural resources with very uneven performance depending on the resource under consideration. On the negative side, increased exploitation rates of forest resources and freshwater resources in some parts of Europe drive the score down. On the opposite end, the indicator showing groundwater bodies in good quantitative status is increasing as a result of a decrease in water abstraction (EEA 2019c), while there has been barely any change in the land area with tolerable soil erosion rates. The remaining environmental functions also show mixed progress with scores that range between 48 and 55.

Europe scores 55 points in sink functions, with relevant differences between global and regional processes. In the global processes, progress is being made in the right direction. On the one hand, the commitments under the Montreal Protocol and its amendments resulted in Europe meeting the environmental standard already in the past and set it in a sustainable trajectory for the future. When it comes to climate change, Europe reduced its per-capita CO_2 emissions at a rate of 1.5% per year between 2013 and 2018 (Eurostat 2019a), which, although positive, is far from the reduction rates required. In this vein, Europe has committed to be climate neutral by 2050 – 20 years later than the reference year used in SESPI –, yet the current trajectory is not enough to even reach existing policy targets (a 55% reduction in GHG emissions compared to 1990) (EEA 2019c). In the case of regional processes, the progress made in cutting chemical pollution is also quite uneven depending on the ecosystem type. In terrestrial ecosystems, Europe has made considerable progress in reducing ozone pollution (34% of the area in good status in 2012 as opposed to 21% in 2012). Some progress (although insufficient) has also been made with regard to acidification and eutrophication in terrestrial ecosystems (21% in good status in 2017 compared to 17% in 2005). Nonetheless, the implementation of existing policies would only lead to increasing the area in good condition to around 50% in 2030 [13] (Amann et al. 2018). In freshwater systems, the percentage of rivers (in length) that met the pollution standards decreased from 72% to 63% in Europe between 2009 and 2015. This needs to be seen in a wider context, since Europe has made some progress in reducing the concentration of some metals and pesticides in surface water bodies (EEA 2018b). Nonetheless, the presence of some ubiquitous, persistent, bioaccumulative and toxic substances such as mercury and brominated diphenyl ethers in many water bodies explains the failure to meet the environmental standard (EEA 2018b). Also in the context of freshwater systems, the groundwater body area in good chemical status has remained stable at European level between 2009 and 2015. This is partly because the area of groundwater bodies in which nitrate concentration – the most relevant pollutant in Europe – has increased, has been compensated by the area in which it has decreased (EEA 2020d). Overall, the average annual mean concentration in groundwater bodies has remained almost constant since 1992 (EEA 2020d). As with rivers, the percentage of coastal water body area that met the environmental standard also decreased between

[13] This figure includes both terrestrial and freshwater ecosystems and is therefore not fully comparable with the previous figures presented, which only refer to terrestrial ecosystems.

2009 and 2015 (76% vs 71%), confirming that, despite progress in addressing some pollutants, the outlook for chemical pollution in many water systems is grim.

Europe scores 48 points in life support functions, as a result of the limited progress made in freshwater ecosystems and the slight stray from environmental standards in terrestrial and coastal water ecosystems. In terrestrial ecosystems, the percentage of habitats in good status decreased from 18% to 16% between 2012 and 2018. This occurred despite the constant increase in the terrestrial area protected as part of the Natura 2000 network, which suggests that the designation of protected areas does not guarantee an effective ecosystem protection (EEA 2019c). In freshwater ecosystems, there was barely any change in the length of rivers in good ecological status between 2009 and 2015, which stayed at 36% of the river length. In the same period, there was a slight worsening in the case of coastal water area in good ecological status (59% vs 57%). These figures are far from the 100% target for all freshwater bodies (including coastal) defined in the Water Framework Directive (European Parliament and European Council 2000), which was meant to be achieved already in 2015.

Uneven progress can be seen in the indicators related to the human health and welfare functions, where Europe scores 52 points. Considerable progress has been made in the last years in improving outdoor air quality, although this is not sufficient to get to 100% of the population below the environmental standard by 2030. Thus, while in 2012, 11% of the European population was exposed to $PM_{2.5}$ levels below those recommended by WHO, in 2017, the percentage increased to 27%. The full implementation of current policies would lead to a substantial improvement, where around 87% of the population would be expected to meet the environmental standard in 2030, which would be accompanied by a substantial reduction in the number of premature deaths attributed to outdoor air pollution (Amann et al. 2018). Indoor air pollution describes a different picture. While compliance with the environmental standard is much higher (94%), there has barely been any change in recent years. Arguably, these areas deserve less attention except in very specific contexts (e.g. in Eastern Europe, where the use of solid fuels for cooking is more common than in other parts of Europe).

As for other welfare indicators, Europe is making some progress in the quality of bathing sites. Between 2014 and 2019, the percentage of bathing sites that met the environmental standard rose from 86% to 88% (EEA 2020c). Although in the right direction, under this trend, not all the bathing sites would meet the environmental standards in 2030. In the case of natural and mixed World Heritage sites, the percentage of sites in good status barely rose slightly from 32% to 33% between 2017 and 2020 (Osipova et al. 2020). This trend is far from the one needed to move all the sites to good quality status.

The results above show that the progress made towards environmental sustainability differs considerably depending on the topic addressed. If we consider the categories in Figure 28, there are three indicators (16%) that are on a sustainable trajectory, zero that describe good progress, four (21%) that report some progress, six that remained almost constant (32%) and six (32%) that are clearly on an unsustainable path. All in all, it cannot be said that Europe is on an environmentally sustainable trajectory.

The trends presented here are largely consistent with those described in the last European State and Outlook of the Environment Report (SOER) (EEA 2019c). This is hardly surprising, as there is some overlap between SESP indicators and those used in SOER to map the status of environment and human health, and therefore, much of the data used for SESPI has also been used in SOER. In this context, it is important to bear in mind that SOER not only contains a much more comprehensive assessment of trends and outlook, which combines data on trends, modelling results and expert input, but also covers many more indicators. While doing so, SOER reports progress towards policy targets.

While the European SOER represents a more comprehensive assessment of trends and outlook, SESPI brings value added in three aspects. First, SESPI has the potential to simplify the communication of indicator trends for non-specialists that lack the time to read long reports such as SOER or that want to easily identify the areas in which a country performs best or worst. Second, one of the insights provided by the European SOER is whether Europe is on track to meet environmental policy targets. However, policy targets and science-based standards often differ (Kutlar Joss et al. 2017; Doherty et al. 2018; UNEP 2020) and therefore, SESPI provides a complementary and necessary perspective on progress towards environmentally sustainability. Without it, countries risk falling short from implementing the actions needed to tackle environmental degradation. Third, not every country has the capacity and expertise to produce a comprehensive SOER report. In those countries, SESPI represents an easy to implement index that can capture the main trends across those indicators related to the functioning of natural capital.

In this vein, it is relevant to note that the paragraphs above discuss the trends in Europe as a whole. As made clear in sections 4.9 and 5.3, each country has its own story, which SESPI can help narrate.

5.4.3. Uncertainty, limitations and further work

Because the normalised score of SESP indicators depends on indicator trends, understanding the uncertainty introduced by the selection of the base year is critical to properly interpret the index and indicator scores. As shown in the uncertainty analysis, several indicator scores are quite sensitive to the baseline year chosen, although except in limited cases, the score consistently captures the direction in which progress is being made. The lack of longer time series for some indicators prevents reaching more solid conclusions. The uncertainty analysis presented is not only relevant for the interpretation of SESPI scores, but the results should also be considered in other indices that use similar methods (Allen et al. 2020; Eurostat 2020b; Hametner and Kostetckaia 2020), since these do not test the influence of the baseline year chosen in their results. In the context of SESPI, some changes in the calculation of observed trends could minimise the uncertainty. For instance, whenever enough data is available, median values calculated in one-year steps could be used. Alternatively, instead of using single years to calculate observed trends, three-year averages could be used. In this first version of SESPI, the method used by Eurostat (2020b) was used, although assuming a linear instead of an exponential evolution. Results have shown to be sensitive to the selection of a linear or a non-linear method to compute trends (Eurostat 2014a). Likewise, the choice of 2030 as target year has been based on its policy relevance, yet while we move closer to that year, its relevance might decrease. Alternatively, SESPI could be computed for a period of ten years from the present in order to avoid being associated with a specific year.

A second aspect that deserves attention is the difference in data availability between indicators. This is something that has already been alluded to in the previous chapter, but it gains more importance in this context. In principle, the same time gap should be used to compute trends, and ideally, data availability should allow to distinguish between short- and long-term trends. Because the data for SESP indicators is updated at different intervals, it was not possible to use the same time gap for all the indicators. Whenever possible, five-year trends were computed to represent short-term evolution. Likewise, the lack of data or of comparability resulted in the exclusion of the indicators on fish resources and access to green spaces. There are some comparability issues with other indicators such as those reported as part of the Water Framework Directive, which also requires the results to be interpreted carefully.

For these reasons, SESPI should be seen as a proof of concept. Compared to other metrics that measure trends towards the SDGs (Eurostat 2020b; Hametner and Kostetckaia 2020; Sachs et al. 2020), SESPI suffers from some limitations in the data availability and comparability aspects. Especially data availability issues are more evident in SESPI because it contains considerably less indicators that other sustainable development metrics. In this first version of SESPI, this is a necessary trade-off between relevance and data quality when selecting indicators to populate the index. Reducing the update gap of some indicators, using nowcasting methods or using expert input to produce outlooks such as in the case of SOER help mitigate the impact of data availability.

5.5. Conclusions

Most environmental and sustainable development metrics show country performance in a given year. Except for a few exceptions in the past, only recently different metrics have emerged specifically intended to measure progress over time, thereby addressing a commonly overlooked aspect in indicator-based sustainability assessments. All these metrics compare current trends with those required theoretically to achieve the SDG targets and therefore fail to represent environmental sustainability when the SDG targets are not aligned with science-based environmental standards. Thus, countries still lack metrics that can answer a simple question: "are we making progress towards environmental sustainability?".

SESPI addresses this gap by incorporating the temporal dimension into the environmental sustainability assessment of countries, thereby complementing the snapshot perspective given by SESI. At the indicator level, SESPI shows progress (or lack thereof) towards science-based environmental standards by comparing current trends with those needed to meet the environmental standards by a certain date. This information is then aggregated through a five-level structure that considers indicators, topics, sustainability principles and environmental functions in order to generate index scores at higher levels that can be used to provide a simple message around the question above.

The results suggest that the progress made at European level is mixed with noteworthy differences between countries and indicators. In this regard, high scores in SESI do not imply high scores in SESPI and vice versa. In general terms, considerable progress is being made in areas such as outdoor air quality, ODS consumption and groundwater abstraction, while trends in other areas such as the exploitation of forest and surface water resources are more worrying. SESPI can be a complement to the more complex picture shown in more comprehensive reports such as SOER, and can be a useful tool to highlight to decision makers and the general public those environmental issues most in need of attention.

As was the case of SESI, SESPI is also presented as a proof of concept. Nevertheless, SESPI not only embeds all the uncertainties in SESI, but also embeds new ones related to the method chosen to compute trends, the selection of the baseline year and the comparability of some indicators. These aspects warrant a careful interpretation of the results.

6. Strong sustainability and the environmental dimension of the SDGs

6.1. Introduction

Chapters 4 and 5 have presented SESI and SESPI as indices that can be used to monitor the environmental sustainability of countries and the progress made towards it. Being part of the ESGAP framework, SESI and SESPI are conceptually aligned with the concept of strong sustainability. Nevertheless, their conceptual soundness does not necessarily reflect their policy relevance and how well they might fit within the existing indicator initiatives. Without the latter, the ESGAP metrics will hardly have any impact beyond academic circles.

At this point, the potential of the ESGAP metrics to have policy impact is unknown. Previous chapters have described the potential uses of SESI and SESPI and insights they can provide, but it was not elaborated how these metrics can complement already established metrics. Of special importance are the SDGs, which are at the core of the 2030 Agenda for Sustainable Development. Each goal is divided into targets that are monitored through indicators. In total, there are 17 SDGs, 69 targets and 247 indicators, 232 of which are unique (UN 2020). As a whole, the SDGs have become a guiding principle for framing environmental and sustainable development policies worldwide.

The overall adequacy and consistency of the SDGs, targets and indicators has been scrutinised several times (e.g. ICSU and ISSC (2015); Spaiser et al. (2017); Nilsson et al. (2018); McGowan et al. (2019); Dawes (2020)). Overall, the SDGs are considered to provide an adequate policy framework (Hák et al. 2016; Janoušková et al. 2018), although they suffer from relevant shortcomings such as trade-offs between targets, non-quantifiable targets (ICSU and ISSC 2015) and problems related to the choice of indicators (Hák et al. 2016; Janoušková et al. 2018; Mair et al. 2018).

The environmental dimension of the SDGs has also been the subject of specific research. This is one of the areas in which the SDGs have improved the most compared to their predecessor, the Millenium Development Goals (ICSU and ISSC 2015; Ekins and Usubiaga 2019; Elder and Olsen 2019). Nonetheless, the role of the environmental dimension in the SDGs differs between goals, targets and indicators. The goals seem to be arranged around the three-pillar structure of sustainable development, with some goals addressing two or more dimensions. Thus, by the wording, the SDGs on climate action (SDG 13), life below water (SDG 14) and life on land (SDG 15) could be considered as purely environmental, while the SDGs on clean water and sanitation (SDG 6), affordable and clean energy (SDG 7), sustainable cities (SDG 11) and responsible production and consumption (SDG 12) would address the economic and/or social dimension in addition to the environmental one (Elder and Olsen 2019). The underlying targets of each goal, on the other hand, are more integrative in that they generally consider several dimensions of sustainable development at the same time. Thus, several environmental targets can be found under non-environmental goals and vice-versa (Elder and Olsen 2019). Nevertheless, as mentioned earlier, some of these targets have been criticised for not being specific enough and difficult to quantify (ICSU and ISSC 2015).

The selection of indicators to operationalise the SDGs and their targets determines whether the SDGs are fit for purpose when it comes to measuring progress towards sustainable development. In this context, different reports intended to monitor the SDGs have used different indicators and methods ultimately leading to different findings and policy conclusions (Janoušková et al. 2018; Miola and Schiltz 2019; Dickens et al. 2020; Lafortune et al. 2020). From now on, in this chapter the main focus is set on the SDG indicators related to the environment (hereinafter referred to as environmental SDG indicators) and subsets thereof. Elder and Olsen (2019) argued that concerns about the cost and feasibility of data gathering and the need to limit the number of indicators of each target ultimately diluted the environmental content in the SDG indicators, thereby creating a disconnect between some targets and indicators. Ultimately, this can result in a bias towards indicators that measure what it can be easily measured, instead of what should be measured. This bias also affects which topics are prioritised in the decision-making process. As Campbell et al. (2020, p. 448) put it "[w]e use existing data to identify priorities, but priorities for data collection are identified on the basis of which topics are priorities".

A few researchers have explored whether the environmental SDG indicators are fit for purpose in the context of environmental sustainability. Campbell et al. (2020) argued that only about a dozen of the more than 90 environmental SDG indicators measured environmental state and trends, with most indicators focusing on other aspects such as environmental policies, links between people and the environment or sustainable consumption or production patterns. Using a different indicator typology, Dickens et al. (2020) concluded that ecosystem health and biodiversity indicators were insufficiently represented among the indicators intended to monitor the status and trends of natural resources. Based on a quantitative assessment, Zeng et al. (2020) showed that there is limited correlation between the environmental SDG indicators and biophysical indicators of biodiversity conservation, therefore questioning the capacity of the environmental SDG indicators to characterise environmental sustainability (Zeng et al. 2020). This is especially worrying if, as some authors suggest, the environmental dimension of the SDGs is already downplayed in comparison to the economic and social dimensions (Neumann et al. 2017; Eisenmenger et al. 2020).

In this regard, in chapter 1, it was argued that, as a set, the SDG indicators did not adequately represent environmental sustainability. While this remains true as a general statement, it deserves a more detailed analysis that sheds light on the differences between the ESGAP metrics and the SDG indicators. Only by understanding these differences, the value added brought by the ESGAP metrics can be made more evident. Thus, the goal of this chapter is to assess the suitability of the SDG indicator sets to measure environmental sustainability in relation to different environmental and resource topics, and to identify overlaps and complementarities between the SDGs and the ESGAP metrics. While doing so, it responds to the third research question of this thesis: *Are the ESGAP metrics complementary to SDG-based metrics?* Given the existence of various SDG indicator-based assessments, different indicator sets are considered (OECD 2019; Eurostat 2020b; Sachs et al. 2020; UN 2020). The assessment is carried out in two stages that help navigate the SDG framework from its three-pillar structure to the environmental dimension and then to the environmental sustainability features, where the ESGAP metrics reside. Thus, the assessment first identifies the SDG indicators that are related to the environmental dimension using the rationale used by Campbell et al. (2020). Then it interrogates those indicators using the criteria of strong sustainability indicators proposed in the introduction and further elaborated in the next section. In order to understand the links between the ESGAP metrics and the SDGs, the structure of SESI (functions, sustainability principles, topics) is used to identify gaps during the assessment. The main novelty of this exercise relies on the proposition and use of specific strong sustainability criteria to interrogate the environmental SDG indicators. This provides a more analytical perspective on their potential to monitor environmental sustainability. The assessment is used to describe the main differences between the environmental SDG indicators and the ESGAP metrics, and the value added brought by the latter.

The chapter is organised as follows. Section 6.2 explains the criteria environmental sustainability indicators need to meet. Section 6.3 describes the methodology followed in the qualitative assessment of the SDG sets, while section 6.4 presents the results. Sections 6.5 and 6.6 discuss the main findings and conclude.

6.2. Strong environmental sustainability indicators in the context of the SDGs

The ESGAP framework argues that there is limited substitution capacity between natural capital and other types of capital due to the inability of non-natural capital to fulfil several environmental functions of natural capital. Within natural capital itself, the functions provided by specific elements cannot be commonly replaced by those provided by other elements either. Thus, from a strong sustainability perspective, development should ensure that the unique functions provided by natural capital are sustained over time, irrespective of those of manufactured, social and human capital (Ekins et al. 2003a).

The suitability of the SDG indicators to reflect strong sustainability has previously been assessed from two perspectives: the structure of the indicator sets or indices and the phenomena they describe. The former is related to the issue of substitutability between the functions provided by different types of capital and between the diverse environmental functions provided by natural capital. The latter, on the other hand, is related to whether the individual indicators reflect the environmental functions of natural capital or describe unrelated phenomena.

Regarding the structure, Rickels et al. (2016) argued that the SDG indicators, being an indicator set without explicit treatment of trade-offs, could be considered to represent strong sustainability if one strictly interprets that sustainable development requires all the indicators to be maintained at least at their current level. Similarly, Neumann et al. (2017) claimed that strong sustainability should be implemented through a constancy of natural capital rule. If the constancy of natural capital rule were ignored and the SDG indicators were to be used to compute an index, the elasticity of substitution assumed at the different levels of the index would be the key factor determining the position in the weak-strong sustainability continuum (Rickels et al. 2016).

Other authors put more emphasis on the phenomena that individual indicators describe, thereby assuming that not all the indicators can be used as strong sustainability indicators, independent from the substitution capacity assumed between them. When applying the strong sustainability paradigm in the environmental dimension of sustainable development, Giannetti et al. (2015) argued that only biophysical indicators should be used. Eisenmenger et al. (2020) specified that only biophysical indicators expressed in absolute terms can monitor the transgression of environmental standards such as planetary boundaries, thereby automatically discarding indicators expressed as percentages, ratios or intensities.

The arguments of the previous paragraphs provide pieces of the puzzle, but there are several caveats that make them insufficient on their own to show the whole picture. When it comes to the structure, the choices made during the weighting and aggregation process determine whether an index is closer to the weak or to the strong sustainability proposition. The need to maintain natural capital compared to a baseline is also presented as a precondition, but what should be maintained and at what level is not specified. This is dependent on the concept of environmental sustainability used, which we define as the maintenance of the environmental functions, and hence the maintenance of the capacity of the capital stock to provide those functions over time (Ekins et al. 2003b). In view of these observations, the following criteria for environmental sustainability indicators have been adopted (c.f. chapter 1).

- First, they need to be indicators linked to the environmental functions of natural capital (source, sink, life support, and human health and welfare). Specifically, they should be indicators (or proxies) of environmental pressure, state or impact in most cases, except in the case of human health and welfare functions, where social state indicators would be most appropriate.
- Second, an appropriate reference value is required against which performance can be measured. That reference value should be defined through science-based environmental standards that ultimately represent the conditions under which the functioning of natural capital is not altered in a way that threatens its capacity to provide ecosystem services in the long-term.
- Third, the indicator needs to be relevant at the national level, given the scope of this thesis.

These criteria conflict with that of Eisenmenger et al. (2020) to some extent. For instance, while intensity indicators that use GDP as nominator should be excluded, their statement is not correct on two grounds. First, absolute pressure indicators obscure spatial disparities, which are key not only for the environmental functions described above, but also for some of the planetary boundaries (e.g. water, biosphere integrity and land system change). Second, when an absolute pressure indicator is compared against an environmental standard, it becomes a percentage or a ratio. Thus, the suitability of ratio indicators does not depend on the format of the indicator, but on the reference used to contextualise its meaning. This reference value can be part of the indicator or included in the normalisation process of an index.

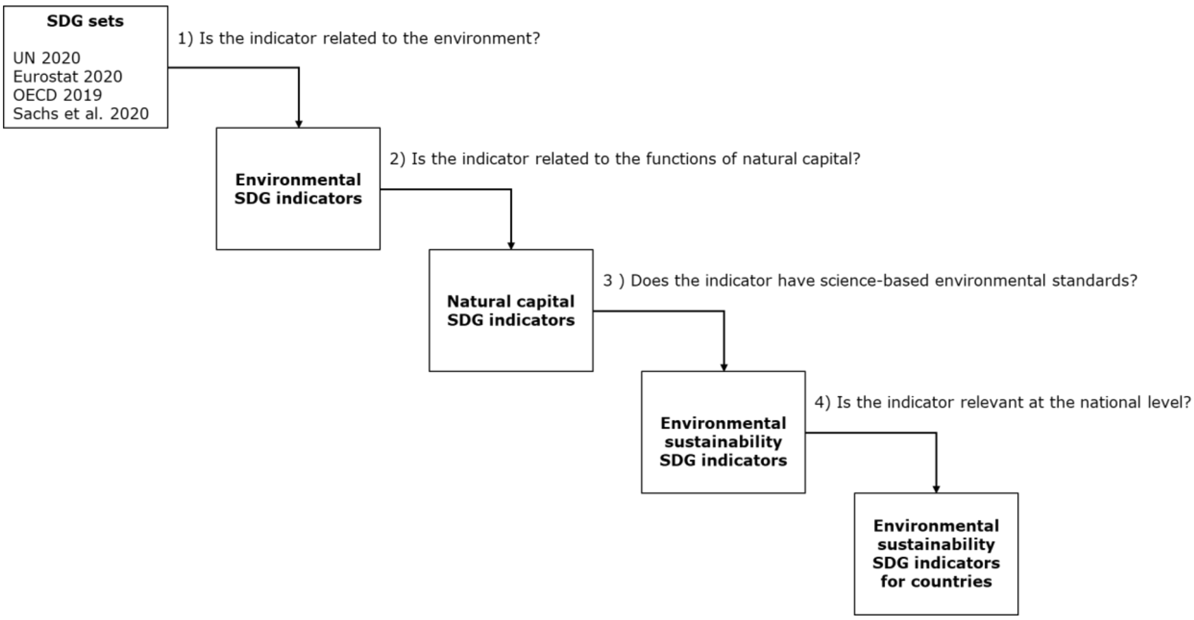
Given that the SDG indicators comprise social, economic and environmental indicators, a first step requires identifying which indicators should be interrogated based on the criteria above. While there is no official classification of SDG indicators across the dimensions they cover, UNEP (2019) and Elder and Olsen (2019) identified which ones are related to the environment. Building on the 93 environmental SDG indicator list proposed by UNEP, Campbell et al. (2020) developed a typology that groups them in the following categories: (1) indicators related to environmental state and trends; (2) indicators related to behaviour or consumption or production patterns; (3) indicators representing linkages between people and the environment; and (4) indicators related to an enabling environment, policy or other mechanisms. Understanding how this typology fits within the DPSIR framework (Drivers-Pressures-State-Impact-Response) (EEA 1999), which provides a useful entry point to assess the suitability of environmental SDG indicators as indicators of strong sustainability. Broadly speaking, the first category includes state indicators or temporal evolutions thereof. The second category includes, among others, pressure indicators. The third one includes social state indicators related to the environment, which can also be considered impacts. The fourth covers responses. There are exceptions to these rules though. Arguably, pressure, state and impact indicators are more suited to monitor environmental sustainability from a strong sustainability perspective, although response indicators can sometimes be used as proxies.

6.3. Methodology

In this report, a qualitative analysis of the suitability of the environmental SDG indicators to monitor environmental sustainability is undertaken. The approach is similar to that of Lafortune and Schmidt-Traub (2019), who analysed the robustness and fitness of SDG monitoring in Europe, in that a set of indicators is assessed against specific criteria.

The analysis starts from the official list of SDG indicators (UN 2020) and sets of SDG indicators used in well-known international SDG assessments. The latter includes Eurostat's 'Sustainable development in the European Union' report (Eurostat 2020b), the OECD's 'Measuring Distance to the SDG Targets 2019' report (OECD 2019) and the 2020 version of the SDG Index (Sachs et al. 2020). In practice, there are many more reports from which indicator sets could have been extracted. After all, countries are expected to adapt the SDG targets and indicators to their national context as reflected in Voluntary National Reviews (Dickens et al. 2019; Lafortune et al. 2020). Nonetheless, the use of four different sets, especially considering the influence of the institutions behind them, is considered enough for the purposes of this exercise. The methodology is arranged in four steps that resemble a decision tree (see Figure 32).

Figure 32: Decision tree used to identify suitable environmental sustainability indicators



In a first step, the environmental SDG indicators in those sets are identified following the indications of Campbell et al. (2020), who did the same with a previous version of the official SDG indicator list. This allows discarding purely social and economic indicators.

In a second step, the environmental SDG indicators that are related to the environmental functions used in the ESGAP framework (namely source, sink, life support, and human health and welfare) are selected and mapped to the structure of SESI (c.f. chapter 4). This step allows discarding environmental SDG indicators that have limited value to assess environmental sustainability from a strong sustainability perspective. While doing so, it sheds some light on the actual weight natural capital has within the environmental dimension of the SDGs. The selection is subjective and includes not only indicators that fit within the topics, but also others that can be considered proxies for the ideal indicator. In this context, it should be noted that some environmental SDG indicators can be allocated to different environmental topics as originally devised in SESI.

In a third step, the remaining natural capital SDG indicators are assessed against the criterion of using science-based environmental standards as reference values to measure environmental sustainability performance. Reference values are used in various ways and therefore, the evaluation of whether these are science-based is context specific. For instance, distance-to-target assessment only considers the reference value to be reached. Nonetheless, when indicators are aggregated in an index, data is usually normalised in a range of 0-100, which requires two reference values to be used as upper and lower bounds as explained in chapter 4. Those reference values can be defined through a value chosen by the indicator producer (e.g. a science-based standard, a policy target, etc.) or through a value based on the sample distribution (e.g. best or worst performer). All these choices determine whether a reference value can be considered to be science-based. Given the characteristics of each indicator set and how they have been used, the approach summarised in Table 34 has been adopted.

To date, the environmental SDG indicators from the official list have been used by the UNEP in the Measuring Progress report series (UNEP 2019, 2021), although considerable data gaps still remain (UNEP 2021). Since the reports do not clarify which SDG targets have quantitative reference values, the wording of the SDG targets and the indicator metadata have been used to assess whether the indicators identified as being related to the functions of natural capital have science-based environmental standards. The SDG Index (Sachs et al. 2020) normalises indicators before aggregating them across goals into a final score. To that end, it uses upper and lower bounds similar to the method used in SESI. The rationale to select those upper and lower bounds is considered to assess their adequacy as environmental standards. The case of OECD (2019) is slightly different. The OECD also normalises the indicator data, but instead of normalising the actual data, it first calculates the distance to a reference value and then normalises the distance values based on the standard deviation of the sample. In contrast to the SDG Index, only the rationale of one reference value needs to be assessed in this case. Last, Eurostat (2020b) does not measure country performance, but trends, similar to SESPI. Nevertheless, trends are, whenever possible, measured against a reference value that defines whether Eurostat measures progress towards environmental sustainability or something else. This requires assessing the rationale of those reference values.

Table 34: Approach used to assess the suitability of reference values

Set	Use	Source
UN (2020)	Contextualise performance	Target description and/or indicator metadata
Sachs et al. (2020)	Normalises country performance	Upper and lower bounds used for normalisation
OECD (2019)	Normalises distance to target values	Reference value against which performance is measured
Eurostat (2020b)	Normalises data on current path vs desired path	Reference value against which trends are measured

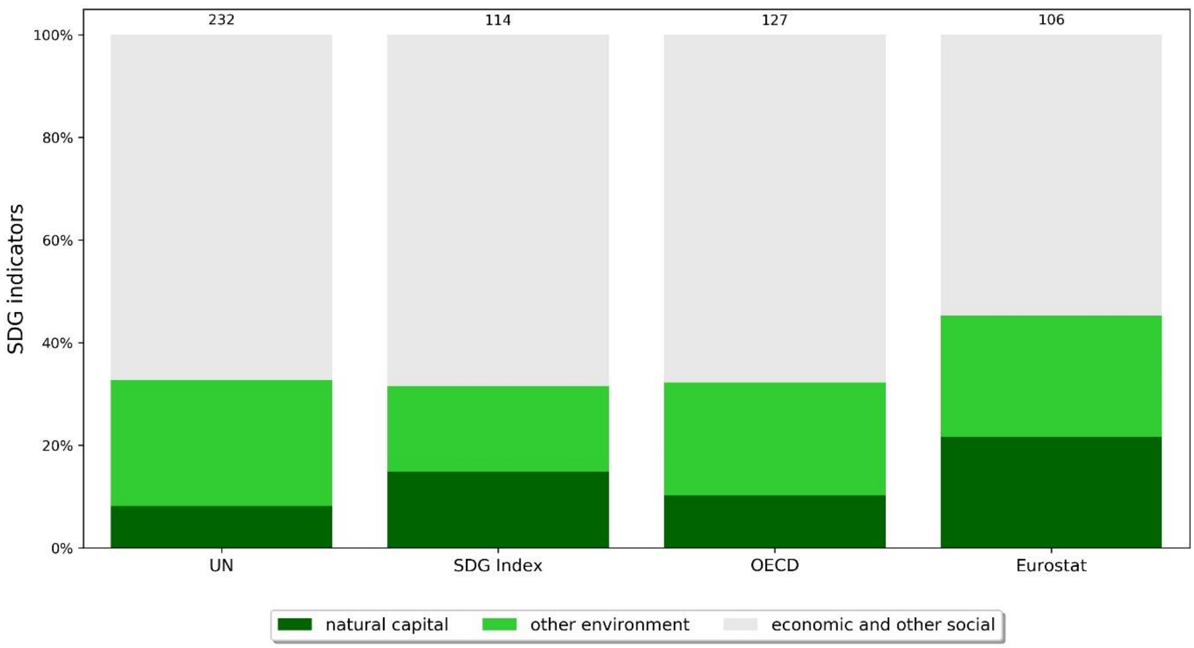
In a last step, the indicators are assessed against their geographical scope. As argued previously, the ESGAP metrics are defined at the national level.

6.4. Results

6.4.1. Environmental and natural capital indicators

Campbell et al. (2020) identified 93 environmental SDG indicators. Using the updated version of the SDG indicator list, the number of environmental SDG indicators decreases to 90, 76 of which are unique. In other sets, the number of environmental indicators is smaller, although similar in relative terms (Figure 33). Thus, the percentage of environmental indicators in the assessed indicator sets ranges between 32% and 45%. The percentage of indicators related to the functions of natural capital (hereinafter natural capital SDG indicators) is much smaller (8-15%), except in the case of the Eurostat indicator set, where they represent 22% of all indicators.

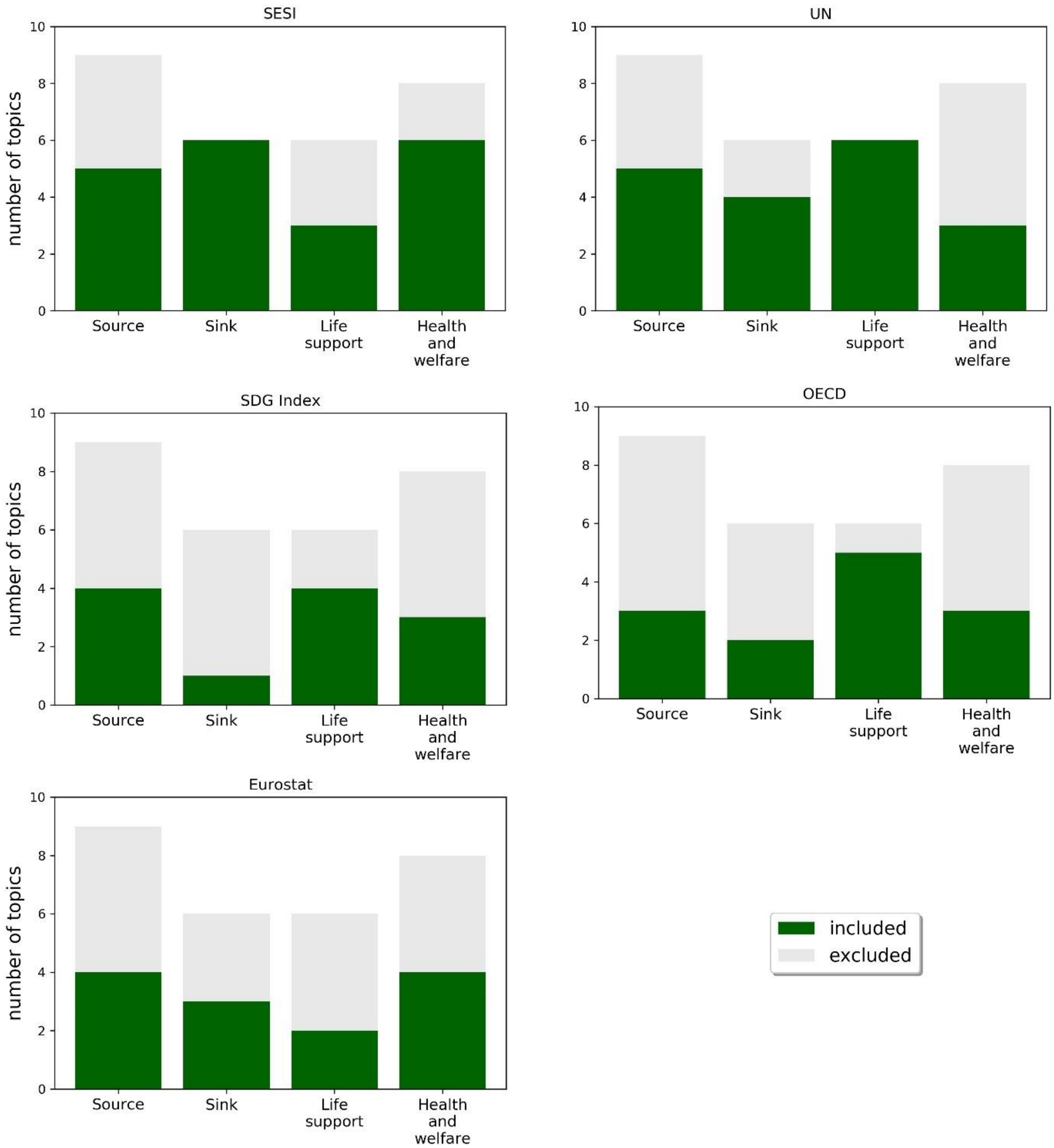
Figure 33: Typology of SDG indicators



The number on top of the bars indicates the number of unique indicators in each set.

Table 35 shows the environmental and resource areas where the SES and natural capital SDG indicators best fit (see Table 43 in Annex 3 for the full list of indicators for each topic). Figure 34 summarises this information by environmental function. SES indicators address most of the topics in the table. This is not surprising, given that the table has been arranged around the structure of SESI. The main topics missing from SESI are exploitation of abiotic raw materials, biodiversity indicators in the life support function and some aspects of health impacts such as noise and chemical pollution. In this vein, it is relevant to note that the list of topics covered in the table is not exhaustive, but rather adapted to the context of the assessment.

Figure 34: Topics covered by SES and natural capital SDG indicators



The coverage of the topics addressed by the natural capital SDG indicators differs depending on the indicator set. In the source function, all the indicator sets fail to cover the extraction of abiotic raw materials. In the case of the SDG Index and the OECD indicators, soil resources are not covered either. This results in non-renewable resources being omitted completely. It is worth noting that the use of abiotic material, as opposed to extraction, is considered in several SDG sets through indicators on material footprints and similar. Nevertheless, these do not address the source function, since they are used as a broad proxy for environmental pressures on the environment. In the case of renewable resources, food, fish and groundwater resources are only covered only in some sets.

Regarding sink functions, none of the SDG indicator sets considers the depletion of the ozone layer and chemical pollution in terrestrial ecosystems. The SDG Index is the set that performs the worst in this category, since it neglects chemical pollution in freshwater and coastal ecosystems as well. In the case of the OECD indicators, it is surprising that climate change is not included. This is perhaps due to the absence of any indicator on GHG emissions in the official UN SDG set at the time of the publication of the OECD report.

In the case of life support functions, all the SDG indicators sets include to varying degrees ecosystem health and biodiversity indicators related to terrestrial, freshwater and marine ecosystems. For biodiversity, this usually occurs through the Red List Index, except in Eurostat, where other biodiversity indices are used. When it comes to ecosystem health, the sets that consider terrestrial ecosystems do so through an index of green vegetation in mountains. Freshwater ecosystems are addressed through changes in extension of ecosystem, while the health of marine ecosystems is reflected by acidity indicators in most cases.

Last, with regard to human health and welfare functions, all the sets provide a good coverage of human health indicators. In the case of Eurostat, it omits indoor air pollution (usually reflected through the use of clean fuels within the household) and drinking water pollution, most likely because Europe has been largely complying with the environmental standards for many years. Instead, it considers noise and chemical pollution, which are not included in any other SDG indicator set. In contrast to human health indicators, none of the indicator sets consider other welfare function indicators properly. Only Eurostat includes one indicator on bathing water quality.

All in all, neither the SES indicators, nor any of the SDG indicator sets covers all the topics represented in Table 35.

Table 35: Environmental and resource areas covered by the environmental SDG and SES indicators

Function	Principle	Topic	Subtopic	SESI	UN	SDG Index	OECD	Eurostat
Source	Renew renewable resources	Biomass	Food resources					
			Forest resources					
			Fish resources					
		Freshwater	Surface water resources					
			Groundwater resources					
	Use non-renewables prudently	Soil	Soil resources					
		Abiotic raw materials	Fossil fuels					
			Metal ores					
			Non-metallic minerals					
Sink	Prevent global warming, ozone depletion	Earth System	Climate change					
			Stratospheric ozone depletion					
	Respect critical levels and loads for ecosystems	Terrestrial ecosystems	Terrestrial pollution					
		Freshwater ecosystems	Surface water pollution					
			Groundwater pollution					
		Marine ecosystems	Marine pollution					
Life support	Maintain biodiversity and ecosystem health	Terrestrial ecosystems	Ecosystem health					
			Biodiversity					
		Freshwater ecosystems	Ecosystem health					
			Biodiversity					
		Marine ecosystems	Ecosystem health					
			Biodiversity					
Human health and welfare	Respect standards for human health	Human health	Outdoor air pollution					
			Indoor air pollution					
			Noise pollution					
			Drinking water pollution					

			Chemical pollution					
	Conserve landscape and amenity	Other welfare	Bathing waters					
			Green areas					
			Natural sites					

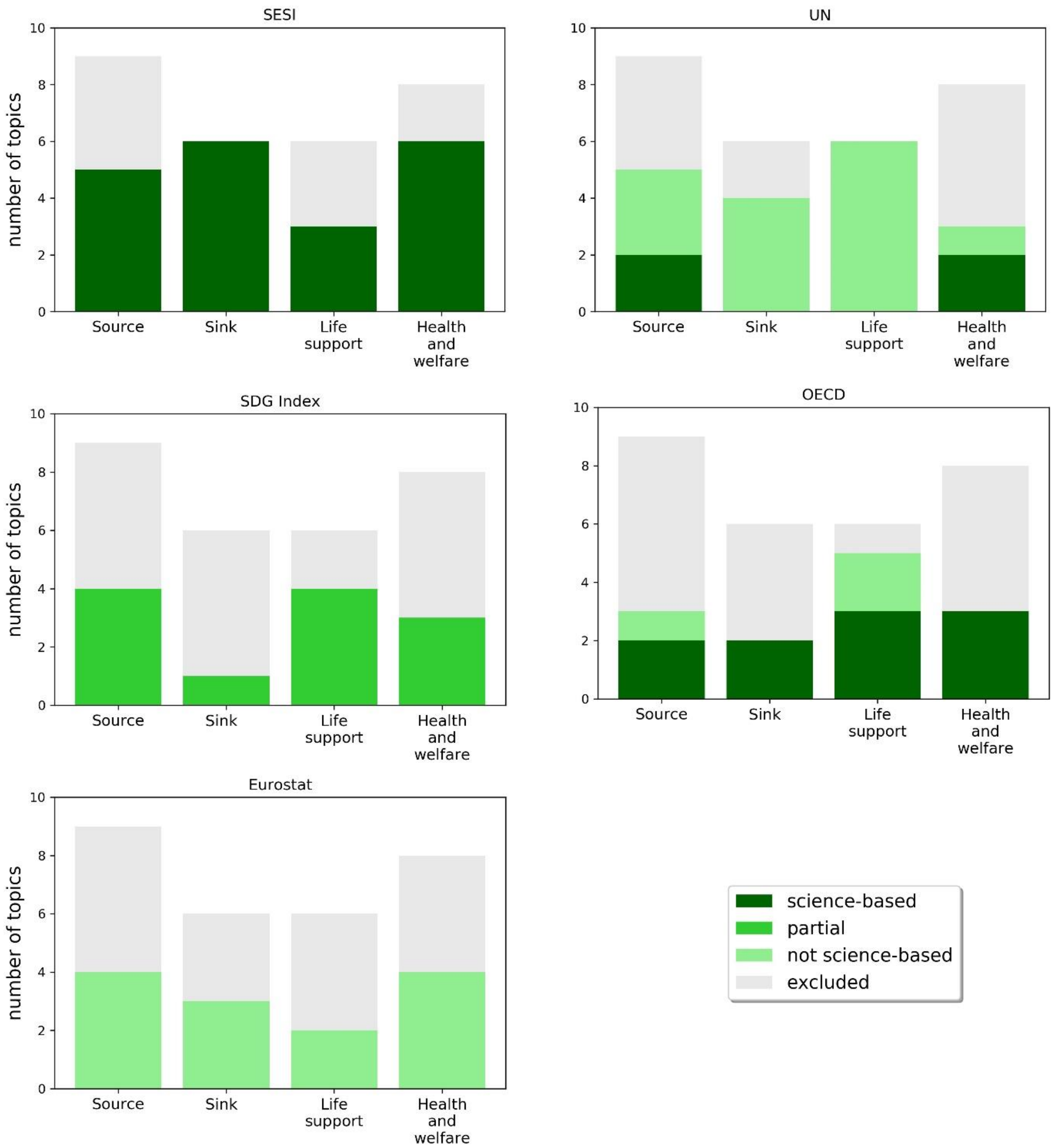
Green shading indicates the availability of indicators for those topics; grey shading indicates the absence of such indicators for those topics.

An equivalent table with the full indicator names is given in Table 43 in Annex 3.

6.4.2. Environmental sustainability indicators

A closer examination reveals limitations of the natural capital SDG indicators in the context of strong sustainability. Figure 35 represents the topics for which there is at least one indicator with a science-based environmental standard. Table 36 shows the same information for topics, instead of for functions (the full information is shown in Table 44 in Annex 3). In the case of SESI, all the topics covered have an indicator with an environmental standard, which results in the adequate coverage of around 70% of the topics in Table 35. This percentage ranges from 0% in the case of the Eurostat set and the SDG Index to 34% for the OECD set. Relevant caveats are described in the following lines, especially in the case of the SDG Index.

Figure 35: Topics with at least one indicator with science-based environmental standards



6.4.2.1. UN SDG Indicators

The UN indicator set has been analysed based on the wording of the SDG targets and the indicator metadata. Most of the targets monitored through the UN SDG indicators related to the functions of natural capital are directional and therefore not quantifiable or specific enough to be used as reference values, let alone as environmental standards. Out of the 19 indicators, only six have quantitative targets. Out of the four that can be considered science-based, two are proxies for health impacts related to drinking water (6.1.1 on access to safely managed drinking water) and indoor air pollution (7.1.2 on reliance on clean fuels and technology), while the other two are indicators of natural capital (6.4.2 on water stress and 14.4.1 on fish stocks). Considering these four indicators, only 14% of the topics in Table 35 are adequately covered by the UN SDG indicator set. None of these topics belongs to the sink and life support functions.

6.4.2.2. SDG Index

The SDG Index is the most nuanced set. It contains 17 natural capital indicators that have been mapped to 12 topics. As explained in the methodology, the SDG Index normalises the underlying indicators using upper and lower bounds similar to the method used in SESI. The upper bound, which represents a normalised score of 100, is chosen through a decision tree that prioritises the following options (Sachs et al. 2020):

- Use absolute quantitative thresholds in SDGs and targets.
- Where no explicit SDG target is available, apply the principle of "leave no one behind" to universal access or zero deprivation.
- Where science-based targets exist, set these as upper bound.
- Where several countries already over-comply with an SDG target, use the average of the top five performers.
- For all other indicators, use the average of the top performers (usually five).

On the other hand, the lower bound, which represents a normalised score of zero, is chosen based on the average of the worst performing countries (in this case, the 2.5th percentile).

The interpretation of the strong sustainability concept in this thesis requires both the upper and lower bounds used in the normalisation process to have a scientific rationale so that the normalised scores can be properly contextualised. In practice, out of the 17 natural capital indicators in the SDG Index, two use the SDG targets as upper bound, two apply the 'leave no one behind' principle, nine use the technical optimum (usually zero pollution, maximum possible score or similar) and four use the average of the best performers (see Table 44 in Annex 3 for more details). In this case, the two that use the SDG targets as upper bound can be considered to be science-based (zero mortality related to air pollution and zero deforestation). To these can be added to the indicators using the 'leave no one behind' and technical optimum values as references. In the case of the indicators using best performers as references, in two of them, the performance of the frontrunners is equivalent to zero pollution or zero pressure. In the other two, the upper bound cannot be considered to have a scientific rationale. While most indicators use an adequate upper bound, none of them uses science-based environmental standards to define the lower bound in the normalisation process.

The effect of using best and worst performers in the normalisation process, as opposed to environmental standards, differs depending on the indicator. In practice, it means that country scores depend on the distribution of the sample, which results in counterintuitive results such as countries that emit seven tonnes of energy-related CO_2 per capita getting a normalised score of 70. The seven-tonnes figure is far from the values usually cited as sustainable (2.0-2.5 tonnes per capita in Akenji et al. (2016), 0.5-2.5 tonnes per capita in chapter 4). After all, best performers can perform better or worse than the environmental standards. Annual mean concentration of $PM_{2.5}$ is a good example of the former. In this case, the average of the five best performing countries is 6.3 μg/m^3, while the guideline value proposed by the World Health Organization is 10 μg/m^3. At the same time, the 2.5^{th} percentile is 87 μg/m^3, far lower than the more lenient interim target proposed by the World Health Organization, which is 35 μg/m^3. In this case, a country with an average annual $PM_{2.5}$ concentrations of 20 μg/m^3 (twice as much as the guideline value recommended by the World Health Organisation) would get a normalised score of 80. This becomes even more problematic if the best and worst performances change every year, as a result of which, a country with the same ambient air pollution levels in two years would have a different normalised score depending on the progress (or lack thereof) made by its peers. The differences between normalising with environmental standards and best/worst performers are also shown in the uncertainty analysis in section 4.9.5.

In total, 15 indicators in the SDG Index partially meet the criterion of having science-based environmental standards, while two do not. When looking at the topics, the 12 topics covered in the SDG Index have at least one indicator with a reference value that can be partially considered to be science-based. This translates as 41% of all the topics being represented by at least one indicator that is partially covered by adequate environmental standards. When considering that both reference values (upper and lower bound) need a scientific rationale, none of the topics would have suitable indicators.

6.4.2.3. OECD SDG Indicators

As explained before, the normalisation process of the OECD indicators depends on the distance to a reference value. Thus, instead of normalising the indicator data, as in the case of the SDG Index, the distance is first calculated and then normalised based using the average performance as reference. As a result, the assessment of the OECD natural capital indicators is based on the adequacy of the reference value used to calculate the distance. The OECD also uses a decision tree to select the reference values (OECD 2019):

- Wherever possible, target levels explicitly specified in the 2030 Agenda are used.
- Where no target value is identified in the text of the 2030 Agenda, target levels are drawn from other international agreements or based on OECD expert judgment.
- If no target value could be identified based on the first two options, the target level is specified based on the best performing OECD countries (90^{th} percentile).
- Finally, for indicators lacking a clear normative direction, the indicator is discarded.

Out of the 13 OECD natural capital indicators, four use SDG targets as references, six use other references, and three use best performances. The four SDG targets are based on the zero mortality and 'leave no one behind' principles, and are therefore considered to be environmental standards. In those using alternative reference values, five seem to have a scientific rationale, while one does not. Last, the best performances used in the last three indicators do not seem to be supported by a scientific rationale. All in all, the percentage of topics covered by at least one adequate indicator decreases from 45% to 34% after considering the use of environmental standards.

6.4.2.4. Eurostat SDG Indicators

Last, Eurostat measures the trends towards a target, rather than performance at a given point in time, as in the previous sets. If available, it uses a quantitative target. Out of the 23 natural capital indicators, only one uses a quantitative target, which is not aligned with science-based environmental standards. As a result, the Eurostat natural capital indicator set does not adequately cover any topic from a strong sustainability perspective.

Table 36: Science-based standards in natural capital SDG and SES indicators

Function	Principle	Topic	Subtopic	SESI	UN	SDG Index	OECD	Eurostat
Source	Renew renewable resources	Biomass	Food resources					
			Forest resources					
			Fish resources					
		Freshwater	Surface water resources					
			Groundwater resources					
	Use non-renewables prudently	Soil	Soil resources					
		Abiotic raw materials	Fossil fuels					
			Metal ores					
			Non-metallic minerals					
Sink	Prevent global warming, ozone depletion	Earth System	Climate change					
			Stratospheric ozone depletion					
	Respect critical levels and loads for ecosystems	Terrestrial ecosystems	Terrestrial pollution					
		Freshwater ecosystems	Surface water pollution					
			Groundwater pollution					
		Marine ecosystems	Marine pollution					
Life support	Maintain biodiversity and ecosystem health	Terrestrial ecosystems	Ecosystem health					
			Biodiversity					
		Freshwater ecosystems	Ecosystem health					
			Biodiversity					
		Marine ecosystems	Ecosystem health					
			Biodiversity					
Human health and welfare	Respect standards for human health	Human health	Outdoor air pollution					
			Indoor air pollution					
			Noise pollution					
			Drinking water pollution					

			Chemical pollution					
	Conserve landscape and amenity	Other welfare	Bathing waters					
			Green areas					
			Natural sites					

Green shading indicates the availability of indicators with environmental standards for those topics; yellow shading means that the reference value of the indicator can be partially considered science-based. Red shading shows that an indicator exists for that topic, but that it does not have an environmental standard. Grey shading indicates the absence of an indicator for those topics.

If a topic has more than one indicator, only the colour of the indicator that is closest to having environmental standards is shown.

6.4.3. National environmental sustainability indicators

All the indicators in the list are relevant at the national level. Thus, the results are not affected by this criterion.

6.5. Discussion

6.5.1. The environmental sustainability dimension of the SDGs

The SDGs are structured around the three pillars of sustainable development. The underlying targets and indicators often address one or more pillars of sustainable development.

Besides the official UN list of SDG indicators, several additional indicator sets have emerged with the intention of monitoring the status of and progress towards the SDGs (OECD 2019; Eurostat 2020b; Sachs et al. 2020). In the four SDG indicator sets, between 32% and 45% of the indicators have an environmental focus. Nonetheless, most of these indicators represent mechanisms intended to address environmental problems, aspects of production and consumption systems, and links between humans and the environment. As a result, only 8-22% of the indicators have a focus on natural capital or its functions, with large differences between the indicator sets. A more detailed look at the topics addressed reveals additional patterns. Generally, SDG indicators tend to cover less topics in the source, sink and human health and welfare functions than SESI, while some SDG sets have more indicators of life support functions. Nonetheless, the last point has relevant caveats. As shown by Dickens et al. (2020), who highlighted several environmental topics missing in the official UN SDG indicator list, ecosystem health indicators are largely missing. This is also the case in the other SDG sets, although with a few exceptions for particular ecosystem types. Regarding biodiversity indicators, some SDG sets have one indicator exclusively on terrestrial biodiversity. Existing biodiversity indicators were not found suitable for SESI.

The focus on natural capital is only one of the criteria that an indicator has to meet to be suitable for monitoring environmental sustainability in the context of strong sustainability. Besides the thematic focus, an indicator needs to have environmental standards against which performance can be measured, and, in the context of this thesis, to be relevant at the national level. Since all the indicators in the SDG sets are relevant at the national level, the second criterion is key to evaluate their adequacy to monitor environmental sustainability.

There are three main reasons for indicators to fail this criterion. First, some indicators lack any quantitative reference value either because the SDG targets were directional or because they were not specific enough. In such cases, it is not possible to check whether the SDG target, or environmental sustainability conditions are ultimately met. Instead, performance can be judged based on the direction of movement of the indicator. For example, increasing forest cover or decreasing annual mean levels of fine particulate matter in cities would be considered to be moving towards sustainable development (Huan et al. 2019). This case only affects the UN SDG indicator set. The second case refers to the indicators that have a quantitative reference value that is not representative of the conditions under which the maintenance of environmental functions is not threatened.

Indeed, many internationally agreed environmental targets are focussed on policy rather than environmental targets (Rounsevell et al. 2020). Such targets may be less ambitious and lack scientific integrity (Doherty et al. 2018). Examples of indicators in this group include different criteria adopted to define water quality in freshwater bodies as in UN SDG indicator 6.3.2 (Proportion of bodies of water with good ambient water quality), the use of policy targets as in the Eurostat's GHG emission indicator, or the selection of arbitrary baselines against which performance is measured as in the UN SDG indicator 15.3.1 (Proportion of land that is degraded over total land area). The third case relates to how indicators are normalised when building indices. Methods that define upper and lower bounds to assign scores based on the sample distribution are not indicative of environmental sustainability, unless the performance of the best and worst performers is aligned with relevant science-based environmental standards. A more useful approach in the context of strong sustainability is to define the upper and lower bounds based on scientific criteria. The indicators used in the SDG Index only fulfil this criterion to a certain extent, since the lower bound is defined based on the sample distribution. This results in countries with high scores in areas such as GHG emissions, biodiversity conservation or air pollution where clear unsustainable patterns have been documented even in high-performing countries.

When assessing the SDG indicator set under these lenses, the number of potential environmental sustainability indicators drops substantially. For reference, SESI covers 69% of the topics in Table 36 with natural capital indicators that have environmental standards. In the SDG sets, this percentage varies from 0% to 34%, although in the case of the SDG Index, 41% of the topics have at least one indicator that partially meets the criterion of having environmental standards. Thus, it can be concluded that, as a general rule, the SDG indicators do not reflect environmental sustainability from a strong sustainability perspective.

Despite the insufficient integration of environmental standards, related concepts were relevant in the formulation of goals and targets. As noted by Elder and Olsen (2019), the Planetary Boundaries and Doughnut Economics frameworks informed the SDG formulation and adoption process, but the choice of indicators became a more technocratic process led by (mainly) statisticians where the cost and feasibility of data gathering was prioritised. This highlights the relevance of the indicator selection process and of the theoretical framework in which they are embedded. According to Hák et al. (2016), the format of the SDGs and their targets provides a policy framework that is not fully reflected in the indicators, which can ultimately result in ambiguous or biased messages. Strong sustainability is clear in that the maintenance of environmental functions is non-negotiable and therefore the indicators selected to monitor environmental sustainability should clearly reflect that. The subset of the SDG indicators analysed above does not reflect that in most cases. SESI and, by extension, SESPI do.

6.5.2. Are the ESGAP metrics complementary to SDG-based metrics?

It is not surprising that SESI and SESPI are more suited to monitor environmental sustainability (as defined with strong sustainability criteria) than any of the SDG indicator sets reviewed, since the ESGAP metrics have been specifically designed for that purpose. SDG sets have only a small subset of environmental sustainability indicators that covers less ground than SESI. After all, besides that small subset, the SDG sets contain other indicators of natural capital that are not adequate to monitor environmental sustainability, other environmental indicators that are not linked to the functions of natural capital, and other indicators that address the social and economic dimensions of sustainable development. Thus, the few environmental sustainability indicators available are hidden within a much larger set of indicators in the SDG sets, which makes both the subsets and the whole SDG sets insufficient for monitoring environmental sustainability. This is more evident when the data is aggregated into indices, since the final score is intended to represent the overall situation of the phenomena described in the underlying indicators.

Beyond the focus on environmental sustainability, SESI also provides more clarity and coherence on what is being measured and how it is being measured. This argument is built by comparing SESI and the SDG sets in relation the type of transgression of the environmental standard being measured, the focus on the territory or consumption, and the rationale for the selection of environmental standards.

Whenever possible, SESI contains indicators that represent the spatial extent of the transgression of environmental standards from a territorial perspective. Thus, indicators tend to take the form of percentages of area, population, etc. that meet the environmental standards. The SDG sets represent a mix of extent and severity indicators in many cases, the latter perspective being missing in SESI. While they contain indicators showing the percentage of population with access to drinking water or using clean cooking fuels, others show national averages of water stress or population averages of air pollution. The latter obscure spatial patterns because of how the information is presented, but capture the severity of the transgression. While none is not superior to the other, there is no apparent logic in how the SDGs combine extent and severity indicators.

As with SESI, most environmental sustainability indicators in the SDG sets have a territorial focus. The SDG Index, on the other hand, combines territorial indicators with footprint-type indicators in order to represent spillover effects. In this context, they even calculate an independent spillover index (Sachs et al. 2020). While the relevance of footprints has already been mentioned (c.f. section 4.10.1), mixing territorial and footprint indicators into an index makes it more difficult to interpret the results. Thus, the ESGAP framework favours territorial aspects because they are easier to be influenced by national policies, while recommending the use of footprint indicators to complement the narrative provided by SESI and SESPI.

The last point refers to the rationale behind the reference values used to measure performance. It has been argued before that environmental standards are intended to depict whether the functions of natural capital can be maintained over time. As such, the standards used in SESI are based on a literature review (c.f. chapter 3). Different decision trees have been used in some of the SDG sets. These favour the use of SDG targets, when possible, but consider targets in international agreements and science-based targets as well. Many of the reference values considered to be science-based in the analysis above refer to 'technical optimums' such as zero mortality or zero emissions, which are either based on a very narrow interpretation of the SDG targets or on an easy fix that is not necessarily realistic. Thus, in general, the environmental standards used in SESI are more robust in that they are based on a review of the scientific literature.

All in all, it can be concluded that the ESGAP metrics are superior to the SDG sets in the context of an analysis of environmental sustainability. Thus, SESI and SESPI address a dimension that neither the SDG sets as a whole, nor the subsets of environmental sustainability indicators can adequately address. Given that the ESGAP metrics also offer advantages related to the coverage of topics, and the coherence and robustness of environmental sustainability assessment, there are no obvious benefits from using the subsets of environmental sustainability SDG indicators to try to monitor environmental sustainability. This is a space that SESI and SESPI can cover, thereby providing insights that the SDG sets cannot.

It is worth noting that the analysis has revealed topics not covered in SESI, which could inform a future revision of the index. Nonetheless, some of the SDG indicators were already considered and discarded because of the absence of adequate science-based sustainability values (c.f. section 4.2.2). These include, for instance, specific indicators on the status of biodiversity (c.f. section 3.4.1). In SESI, the status of biodiversity is part of the information considered in the composite indicators of ecosystem condition, which are largely missing in the SDG sets (Dickens et al. 2020).

6.5.3. Does the above hold true for non-European countries?

Although the ESGAP framework can be implemented in any country, the structure of SESI has been defined based on the data availability in Europe, and therefore, the previous analysis needs to be interpreted in that light. Case studies in non-European countries have shown that SESI could not be computed with as many indicators as in this thesis (Otieno et al. 2021; Trung Thang et al. 2021), although this also holds true for many environmental SDG indicators (UNEP 2021). This raises the issue of whether the advantages attributed to the ESGAP metrics remain if the framework were to be implemented in other countries.

Arguably, the conceptual coherence and robustness of the framework are independent from the form SESI and SESPI take in each country. As a result, when data is available for a minimum amount of SES indicators, the ESGAP metrics will provide insights that the SDG sets cannot. Nonetheless, it will require a careful interpretation of the results, especially, if the distribution of indicators across the four broad environmental functions is unbalanced (e.g. if indicators for life support functions are lacking).

In general, beyond the amount of SES indicators included in the analysis, the implementation of the ESGAP framework demands strong sustainability thinking. It requires acknowledging that there are specific elements and functions of natural capital that need to be preserved, and that defining the levels at which those elements and functions need to be maintained is primarily a task for science. Only by integrating this type of thinking in policy making we will be able to measure what matters in the context of environmental sustainability.

6.6. Conclusions

The results of the qualitative assessment above show that although the natural capital SDG indicators address many of the environmental functions and topics in the ESGAP framework, they generally lack science-based environmental standards that would make them suitable for monitoring environmental sustainability from a strong sustainability perspective. After all, under the strong sustainability proposition, the maintenance of environmental functions is non-negotiable.

In the absence of complementary metrics, the SDG indicators risk giving a misleading message about environmental sustainability. This is particularly worrying if such a message conflicts with the scientific evidence that shows widespread environmental degradation in critical areas and the need to act urgently. Initial evidence points towards this hypothesis (Zeng et al. 2020), but a more detailed analysis is needed, ideally by quantitatively assessing the effects using science-based and alternative reference values in relevant natural capital indicators. The qualitative analysis undertaken in this report is an initial step in this direction. This analysis is nonetheless based on how the related SDG indicators have been used so far. Thus, it focuses on whether environmental standards have been incorporated, rather than on whether they could be incorporated.

The SDG indicators are not strong sustainability indicators. They are intended to monitor progress towards the Agenda 2030, which is not only broader, but also different from environmental sustainability. The ESGAP metrics are therefore complementary in that they monitor a different phenomenon. Arguably, the use of environmental standards could be considered in the refinements that SDG indicators undergo annually, but would have a limited impact because of the relatively few natural capital indicators in the SDG sets. Because of this, there is little sense in adapting the SDGs to monitor environmental sustainability, but to use the ESGAP metrics instead. While these strands should advance in parallel for the time being, in the future, the ESGAP framework could inform the adoption of indicator-based strategies, the same way the Planetary Boundaries and the Doughnut Economics frameworks informed the adoption of the SDGs.

7. Conclusions

7.1. Summary of key findings

Metrics, which include individual indicators, indicator sets and indices, fulfil a variety of functions in the policy cycle. From providing information on the state of the environment to monitoring progress towards policy objectives, environmental metrics have become a key part of environmental governance. Given the widespread environmental degradation that is threatening key life support systems, it is imperative to have metrics that can adequately translate this information to different levels, especially the national one, which is the level at which most environmental policies are implemented.

Chapter 1 has shown that countries still lack robust and resonant metrics that allow them to monitor environmental sustainability as well as progress towards it. Environmental sustainability requires the functions of natural capital to be maintained in the long term. Thus, relevant metrics for countries need to be related to natural capital or its functions, have a reference value that is indicative of environmental sustainability conditions and be applicable at the national level. On the one hand, monetary metrics of weak sustainability fail to capture key aspects represented by biophysical indicators of natural capital. On the other hand, existing sustainable development and environmental metrics have significant limitations when it comes to representing environmental sustainability at the national level from a strong sustainability perspective. In this latter group of metrics, some sets and indices have indicators that mostly focus on other issues such as policies or consumption and production patterns. Likewise, while the use of reference values to contextualise country performance is widespread, these values do not necessarily represent environmental sustainability conditions, but other aspects such as policy targets or best performances. The most notable exception is the Planetary Boundaries framework, but this has not yet been implemented convincingly at the national level.

This research gap has been addressed through the ESGAP framework (extensively described in chapter 2), which builds on the concepts of strong sustainability, critical natural capital, environmental functions and science-based environmental standards. Building on the original SGAP approach developed two decades ago, the renewed ESGAP framework has been designed with a stronger focus on implementation. To that end, new indices (SESI and SESPI) have been proposed, and the concept of science-based environmental standards has been made operational across a range of environmental issues. The latter is the main distinguishing factor between ESGAP metrics and other metrics.

The literature review in chapter 3 has provided an overview of existing environmental standards across the four main function categories (source, sink, life support, and human health and welfare) considered in the ESGAP framework. The standards described come from a variety of sources, most prominently peer-reviewed papers, scientific reports and European policy documents. The latter was only chosen when standards characterised the sustainability of specific elements of natural capital and had a clear scientific rationale.

Environmental standards play a central role in the construction of SESI and SESPI, which were described and computed in chapters 4 and 5 for 28 European countries. SESI and

SESPI represent different aspects of environmental sustainability from a strong sustainability perspective. On the one hand, SESI measures, at a given point in time, the performance of countries against environmental standards through 21 indicators that cover a variety of environmental and resource topics arranged around sustainability principles and broad function categories (source, sink, life support, and human health and welfare). Some relevant topics such as food, biodiversity and abiotic resources are insufficiently present in the indicators because of a lack of appropriate environmental standards.

The relatively low SESI scores obtained by the European block (with great variations among countries) presented in chapter 4 suggest that some environmental functions are threatened in Europe. Scores are generally lower in the sink and life support functions compared to source and human health and welfare functions. Given the importance of life support functions in enabling life, these results are specially worrying.

SESPI, on the other hand, measures progress towards or away from environmental standards, thereby complementing the snapshot perspective given by SESI. SESPI shares the general structure of SESI, although because of data quality aspects it only features 19 of the 21 indicators used in the latter. Each indicator in SESPI compares observed trends with those that would be needed to reach its corresponding environmental standards by 2030. The results show mixed progress towards environmental sustainability in Europe with, as in the case of SESI, very uneven performance across countries. The source function has the lowest scores for Europe as a whole, although progress towards environmental sustainability in the other functions is quite limited as well. While these results provide a broad overview of the situation, at the level of the individual indicators diverging trends can be observed, which demands a context-specific interpretation of the results should these be used in policy analyses. Of course, the index results are subject to several caveats, some of the most important being related to the choices made during the construction of the indices. Uncertainty in respect of many of these choices has been presented and therefore, the results should be interpreted taking these uncertainty analyses into account.

When combined, SESI and SESPI have the potential to fill the indicator gap identified in chapter 1. Both indices together with the function scores can provide easy-to-understand messages to politicians, high-level policy makers and the general public around environmental sustainability performance and progress, while the information at lower levels can be used by a more technical audience. The indices and the underlying indicators are not intended to replace existing environmental and sustainability indicator sets, but to complement them by monitoring a specific phenomenon: environmental sustainability seen through the lens of strong sustainability. In this vein, the analysis undertaken in chapter 6 has shown the complementary nature of the ESGAP metrics and the environmental SDG indicators, but also the differences between them. While there is some overlap between the indicators used in SESI and SESPI, and those used in various SDG indicator sets, the latter often lack the environmental standards required under the strong sustainability perspective. This is the niche the ESGAP metrics can fill.

7.2. Research and policy implications

7.2.1. Research implications

Because of the lack of specificity, the concept of sustainable development has been adapted by different users with different (and sometimes conflicting) discourses (Greco et al. 2019). As a result, the conditions for environmental sustainability have remained vague in the general sustainable development narrative. The distinction between weak and strong sustainability with regard to the substitutability of the functions provided by natural capital is a key distinguishing feature of how environmental sustainability is translated into metrics, which ultimately determines how it is reflected in narratives. Although strong sustainability is aligned more closely with the biophysical environment, countries still lack adequate metrics to monitor environmental sustainability from this perspective. The ESGAP framework advances the understanding of environmental sustainability by making it more concrete and defining key criteria that relevant indicators need to meet. One of the key criteria is the existence of science-based environmental standards, for which this thesis provides the most comprehensive overview to date, including the identification of areas in which such standards are missing or insufficiently robust. This is a relevant contribution to the identification of science-based reference values, which is conceptually linked to other frameworks such as Planetary Boundaries (Steffen et al. 2015b), the Science-Based Targets initiative (Walenta 2020), and the Ecosystem Accounting section of the System of Environmental-Economic Accounting (UNDESA 2021). The latter is of special interest, since the manual contains a full chapter on reference values for ecosystem condition, which can potentially overlap with some environmental standards described in chapter 3.

The ESGAP metrics computed in this thesis – SESI and SESPI – advance the measurement of environmental sustainability compared to existing metrics. They closely follow the steps of the OECD manual on composite indicators (OECD and JRC 2008) and provide clear links between the theoretical framework and the choices made during the construction of the indices, something that is not explicitly done in many other cases, since the conceptual framework and other steps are sometimes insufficiently discussed (Kwatra et al. 2020). Chapter 4 even includes uncertainty and statistical analyses in order to understand how normalisation, weighting and aggregation choices affect the results and the statistical coherence of the indices. This is sometimes done through external audits (Papadimitriou et al. 2019; Papadimitriou et al. 2020) or omitted altogether. The results of SESI and SESPI can inform the work of the European Environment Agency in the context of their annual indicator reports or the next state of the environment report to add a strong sustainability dimension to the policy dimension they often adopt. As argued before, there is some overlap between these dimensions in Europe, although relevant differences exist.

SESI and SESPI provide useful information at different levels: from indicators to the final index score. Index scores are particularly relevant in the context of Beyond GDP indicators, since they can be used as headline metrics of environmental sustainability and therefore complement GDP when they become more mature.

Beyond the ESGAP metrics, this thesis also highlights the limitations of the SDG indicators for monitoring environmental sustainability and progress towards it. It does so by identifying the specific criteria strong sustainability indicators need to meet and by interrogating SDG indicators against them. Thus, it adds to the growing evidence of the limitations of the SDGs in this context.

7.2.2. Policy implications

Arguably, Europe is one of the regions with the strongest environmental policy. Its long-term vision is closely aligned with the strong sustainability paradigm (EC 2014b) and, as such, several policies require specific elements of natural capital to be in good condition in order to preserve its capacity to provide ecosystem goods and services. Examples include, but are not limited to, binding environmental targets for freshwater bodies (European Parliament and European Council 2000), relevant terrestrial ecosystems (European Council 1992), ozone layer (European Parliament and Council of the European Union 2009), drinking water (European Council 1998), bathing water (European Parliament and European Council 2006), etc. Nonetheless, some environmental targets are not aligned with science-based standards (e.g. outdoor air quality (Kutlar Joss et al. 2017)), while other areas are missing binding targets altogether (e.g. soils (EEA 2019c) or extraction of fossil fuels). But beyond policies that specifically target environmental concerns, social and economic policies need to be aligned with strong sustainability as well.

In Europe, the European Green Deal (EC 2019b) represents the development strategy for the upcoming years. While sustainability and environmental challenges feature very prominently, it remains a growth strategy (Eckert and Kovalevska 2021). What seems clear is that, generally speaking, previous (sustainable) development strategies in the EU have proven insufficient to improve the state of the environment as shown by EEA (2019c). After all, there are many questionable decisions taken in the name of economic growth that are not compatible with environmental sustainability. A few examples include plans to open a new coal mine and to explore new oil fields in the UK (Ambrose 2021; Willis 2021), the recent opening of a coal-fired power plant in Germany (DW 2020), the lack of action to stop the collapse of Europe's largest saltwater lagoon in Spain (Sánchez 2021) and many more. Of course, this is not restricted to the Europe. International examples include the disregard of Saudi Arabia of the evidence that quantifies unburnable fossil fuel stocks (Smith 2021) or Brazil's plans to develop the Amazon (Woodward 2019). While these are just examples of controversial decisions that do not necessarily express those countries' attitude towards environmental challenges, they clearly violate the spirit of the SDGs, thereby showing that, in practice, economic growth and sustainable development do not have the same policy priority.

The little weight given to scientific evidence in certain decisions has led to an increasing number of calls from scientists (Ripple et al. 2017; Ripple et al. 2020; Wiedmann et al. 2020; Albert et al. 2021), and from civil society such as the Fridays for Future movement, for decision makers to better integrate scientific evidence in the decision-making process. As a result, there is an increasing demand for strong sustainability thinking, and the ESGAP framework can be a useful tool in its promotion and communication. On the one hand, these indices can provide easily digestible information for decision makers to gain an overview of the areas that need attention either because environmental functions are threatened or because trends are going in the wrong direction. This is a unique feature of SESI and SESPI compared to other metrics. Beyond that, the ESGAP framework also highlights the need for national science-based targets. This should become an integral part of the design and monitoring of national policies and international multilateral environmental agreements.

7.3. Outlook

The work presented in this thesis represents the first full attempt to operationalise the ESGAP framework after the conceptual and illustrative work of Ekins and Simon (2001) in the early 2000s. In that paper, the authors measured a policy gap for two countries – rather than a sustainability gap – and computed an index whose structure and method lacked an explicit link to the theoretical framework. Although this thesis represents a considerable step forward in the theoretical conceptualisation and practical implementation of the ESGAP framework, the indices computed can only be considered a proof of concept at this point. Going forward, there are three different work streams that would help consolidate the ESGAP framework in environmental sustainability monitoring and increase the maturity and robustness of the underlying metrics.

7.3.1. Consolidation of SESI and SESPI as headline metrics of environmental sustainability

This first version of the ESGAP metrics has been tailored to the European context and therefore relies on many environmental standards that are specific to Europe and on data that in some cases is only produced as part of European environmental legislation. Because of this, the computation of the two indices might prove more difficult in other countries or not possible at all.

The availability of environmental standards will likely increase in the future as part of the evolution of the Planetary Boundaries framework (e.g. Gleeson et al. (2020)) or the growing interest in science-based targets. Beyond these two frameworks, research on environmental standards can also be structured around the ESGAP framework. Whether this happens through workshops such as in the inception of the Planetary Boundaries framework or through remote expert elicitation processes, there is potential to increase the knowledge base – and therefore the availability – of environmental standards.

Data availability is also one of the main constraints countries will face to compute SESI and SESPI. This has become evident in the countries in which the ESGAP framework is being implemented, namely New Caledonia (Comte et al. 2021), Vietnam (Trung Thang et al. 2021), Kenya (Otieno et al. 2021), the Bahamas, China (UCL 2021a) and Japan (UCL 2021b). These countries are using the methodology developed here as guidance, but the ultimate selection of indicators differs. This hampers the comparability of the results, but it allows the capture of national specificities that could otherwise go unnoticed. As an example, SESI is computed in Vietnam and Kenya using 14 and 12 indicators respectively as opposed to the 21 used in the European version. Relevant insights are being extracted from ongoing pilots. For instance, pilots show the difficulties related to monitoring the status of the elements of natural capital that are linked to the life support functions, yet they highlight the potential of the ESGAP framework to be leveraged as a communication tool and to complement existing indicator initiatives in the pilot countries. Additional pilots would help draw down additional lessons that can be used in the future when revising the methodology of the index.

Since the results of the pilots cannot be compared, an exercise should be devoted to test how many indicators can be computed using international databases alone. Such an exercise would help project SESI and SESPI internationally. A review of the databases is already available (Fairbrass et al. 2020b), but the index has not been computed yet.

7.3.2. Bringing the notion of environmental sustainability into sustainable development narratives

The ESGAP framework and its metrics are conceptually sound, but without an appealing narrative they risk getting lost in what is has been termed an 'indicator zoo' (Pintér et al. 2012). As shown in Table 2, there are several metrics that are related to the environmental dimension of sustainable development. These include, but are not limited to, the SDG indicators (and index), the Planetary Boundaries framework, the Environmental Performance Index and the Ecological Footprint. All provide complementary information, but the sustainable development narrative is currently dominated by the SDGs, although the other metrics also have their niches.

The SDGs represent an unprecedented consensus on the direction of development and, as such, they have been embedded in numerous policies. The actions required to implement the 17 goals are ultimately reflected in their 169 targets and 232 indicators. Nonetheless, the environmental SDG indicators have been found to be problematic to monitor the state of the environment (Campbell et al. 2020; Dickens et al. 2020; Zeng et al. 2020). The qualitative assessment presented in chapter 6 also suggests that environmental SDG indicators can be misleading when it comes to monitoring environmental sustainability from a strong sustainability perspective. Translating the qualitative insights into numbers would make this problem more evident. In this vein, the ESGAP metrics could fill this national level environmental sustainability monitoring gap, in the same way as the Planetary Boundaries framework does at the global level. The inclusion of the ESGAP framework in a recent UNEP report (UNEP 2021) and its reference in the Dasgupta review on the economics of biodiversity (Dasgupta 2021) provides a good starting point to increase the exposure of the ESGAP framework and the promotion of strong sustainability thinking.

Beyond the SDGs, the narrative developed through ESGAP metrics could also be complemented with other metrics to address some of the aspects not covered in this thesis. For instance, ESGAP metrics could be complemented with footprint indicators and severity indicators as long as a suitable environmental standard were to be found. This would allow addressing the consumption and the severity dimensions without the need to integrate these indicators in the indices. Likewise, the narrative could also be expanded with global indicators such as in the Planetary Boundaries framework to incorporate Earth System processes for which no environmental standards exist at the national level.

7.3.3. Development of the monetary environmental sustainability gap

A third metric (not quantified here) was originally proposed by Ekins in the original SGAP approach (Ekins and Simon 1999). The monetary environmental sustainability gap represents the aggregated monetary value of the costs (i.e. abatement, avoidance, maintenance, restoration and protection costs) needed to close the physical sustainability gap (i.e. the gap between sustainability conditions and the SESI) for the relevant elements of natural capital, assuming previous losses are reversible. Some thoughts on how partial- or general-equilibrium estimates of the monetary gap could be produced have been discussed in Ekins (2011), although as in the case of SESI and SESPI, it is likely that actual implementation of the indicator would require further exploration of the practicalities involved. Given the similarities between this indicator and Hueting's 'Sustainable National Income' concept (Hueting and De Boer 2001), the limited attempts to quantify the latter (Gerlagh et al. 2002) would be a good starting point to devise how to quantify the monetary environmental sustainability gap. This would allow moving from the description of the problem – as in SESI and SESPI – towards a solutions-oriented approach and expanding the timeframe of the assessment compared to SESPI.

References

Akenji, L., M. Bengtsson, R. Bleischwitz, A. Tukker, and H. Schandl. 2016. Ossified materialism: introduction to the special volume on absolute reductions in materials throughput and emissions. *Journal of Cleaner Production* 132: 1-12.

Al-Sharrah, G. 2010. Ranking Using the Copeland Score: A Comparison with the Hasse Diagram. *Journal of Chemical Information and Modeling* 50(5): 785-791.

Albert, J. S., G. Destouni, S. M. Duke-Sylvester, A. E. Magurran, T. Oberdorff, R. E. Reis, K. O. Winemiller, and W. J. Ripple. 2021. Scientists' warning to humanity on the freshwater biodiversity crisis. *Ambio* 50(1): 85-94.

Alcamo, J., T. Henrichs, and T. Rösch. 2000. *World water in 2025: Global modeling and scenario analysis for the world commission on water for the 21st century*. A0002. Kassel. Kassel World Water Series.

Allen, C., M. Reid, J. Thwaites, R. Glover, and T. Kestin. 2020. Assessing national progress and priorities for the Sustainable Development Goals (SDGs): experience from Australia. *Sustainability Science* 15(2): 521-538.

Alley, W. M. and S. A. Leake. 2004. The Journey from Safe Yield to Sustainability. *Groundwater* 42(1): 12-16.

Alvarenga, R., I. Lins, and J. Almeida Neto. 2016. Evaluation of Abiotic Resource LCIA Methods. *Resources* 5(1): 13.

Amann, M., M. Anderl, J. Borken-Kleefeld, J. Cofala, C. Heyes, L. Höglund-Isaksson, G. Kiesewetter, Z. Klimont, L. Moosmann, P. Rafaj, R. Sander, W. Schöpp, M. Tista, F. Wagner, and W. Winiwarter. 2018. *Progress towards the achievement of the EU's air quality and emissions objectives*. Laxembourg: International institute for Applied Systems Analysis.

Ambrose, J. 2021. UK faces legal action over North Sea oilfield exploration plans. *The Guardian*, section.

Andersen, I., N. Ishii, T. Brooks, C. Cummis, G. Fonseca, A. Hillers, N. Macfarlane, N. Nakicenovic, K. Moss, J. Rockström, A. Steer, D. Waughray, and C. Zimm. 2020. Defining 'science-based targets'. *National Science Review*.

Arthington, A. H., S. E. Bunn, N. L. Poff, and R. J. Naiman. 2006. THE CHALLENGE OF PROVIDING ENVIRONMENTAL FLOW RULES TO SUSTAIN RIVER ECOSYSTEMS. *Ecological Applications* 16(4): 1311-1318.

Ashmore, M. R. and R. B. Wilson. 1993. *Critical levels of Air Pollutants for Europe. Background Papers prepared for the ECE Workshop on critical levels*. Egham.

Ashmore, M. R. and A. W. Davison. 1996. Towards a critical level of ozone for natural vegetation. In *Critical levels for ozone in Europe: testing and finalising the concepts*, edited by L. Kärenlampi and L. Skärby. Kuopio: United Nations Economic Commission for Europe.

Barbier, E. B. and J. C. Burgess. 2017. Natural Resource Economics, Planetary Boundaries and Strong Sustainability. *Sustainability* 9(10): 1858.

Bass, S. 2009. Keep off the grass. *Nature Reports Climate Change*(0910): 113-114.

Bavington, D. 2010. From Hunting Fish to Managing Populations: Fisheries Science and the Destruction of Newfoundland Cod Fisheries. *Science as Culture* 19(4): 509-528.

Bazzoffi, P. 2008. Soil erosion tolerance and water runoff control: minimum environmental standards. *Regional Environmental Change* 9(3): 169-179.

Becker, W., M. Saisana, P. Paruolo, and I. Vandecasteele. 2017. Weights and importance in composite indicators: Closing the gap. *Ecological Indicators* 80: 12-22.

Becker, W., D. Benavente, M. Dominguez Torreiro, C. Moura, A. Neves, M. Saisana, and D. Vertesy. 2019. *COIN Tool User Guide*. Luxembourg: Joint Research Centre.

Bell, J. N. B. 1993. A reassessment of critical levels for SO2. In *Critical levels of air pollutants for Europe. Background papers prepared for the United Nations Economic Commission for Europe Workshop on Critical levels*, edited by M. R. Ashmore and R. B. Wilson. Egham.

Bell, S. and S. Morse. 2008. *Sustainability Indicators: Measuring the Immeasurable?* Second Edition ed. London and Sterling: Earthscan.

Bell, S. L., C. Phoenix, R. Lovell, and B. W. Wheeler. 2014. Green space, health and wellbeing: making space for individual agency. *Health & Place* 30: 287-292.

Benini, A. 2019. *Priorities and preferences in humanitarian needs assessments - Their measurement by rating and ranking methods*. A note for Okular Analytics.

Beverton, R. J. H. and S. J. Holt. 1957. *On the Dynamics of Exploited Fish Populations*. London: Chapman & Hall.

Biggs, R., G. D. Peterson, and J. C. Rocha. 2018. The Regime Shifts Database: a framework for analyzing regime shifts in social-ecological systems. *Ecology and Society* 23(3).

Blomqvist, L., B. W. Brook, E. C. Ellis, P. M. Kareiva, T. Nordhaus, and M. Shellenberger. 2013a. The Ecological Footprint Remains a Misleading Metric of Global Sustainability. *PLOS Biology* 11(11): e1001702.

Blomqvist, L., B. W. Brook, E. C. Ellis, P. M. Kareiva, T. Nordhaus, and M. Shellenberger. 2013b. Does the Shoe Fit? Real versus Imagined Ecological Footprints. *PLOS Biology* 11(11): e1001700.

Bobbink, R. and J.-P. Hettelingh. 2011. Review and revision of empirical critical loads and dose-response relationships. Paper presented, 23-25 June 2010, Noordwijkerhout.

Bobbink, R., D. Boxman, E. Fremstad, G. Heil, A. Houdijk, and J. Roelofs. 1992. Critical loads for nitrogen eutrophication of terrestrial and wetland ecosystems based upon changes in vegetation and fauna. In *Critical loads for nitrogen*, edited by P. Grennfelt and E. Thörnelöv. Lökeberg Nordic Council of Ministers,.

Bobbink, R., H. Tomassen, M. Weijters, L. v. d. Berg, J. Strengbom, A. N. Sabine Braun, K. Schütz, and J.-P. Hettelingh. 2015. Effects and Empirical Critical Loads of Nitrogen for Europe. In *Critical Loads and Dynamic Risk Assessments: Nitrogen, Acidity and Metals in Terrestrial and Aquatic Ecosystems*, edited by W. d. Vries, et al.: Springer.

Borucke, M., D. Moore, G. Cranston, K. Gracey, K. Iha, J. Larson, E. Lazarus, J. C. Morales, M. Wackernagel, and A. Galli. 2013. Accounting for demand and supply of the biosphere's regenerative capacity: The National Footprint Accounts' underlying methodology and framework. *Ecological Indicators* 24: 518-533.

Brack, W., B. I. Escher, E. Müller, M. Schmitt-Jansen, T. Schulze, J. Slobodnik, and H. Hollert. 2018. Towards a holistic and solution-oriented monitoring of chemical status of European water bodies: how to support the EU strategy for a non-toxic environment? *Environmental Sciences Europe* 30(1): 33.

Bradley, I., J. King, A. Bhogal, P. Bellamy, J. Hollis, S. O'Keefe, and R. Sakrabani. 2004. *Soil organic matter as an indicator of soil health*. London: Department for Environment, Food and Rural Affairs.

Brand, F. 2009. Critical natural capital revisited: Ecological resilience and sustainable development. *Ecological Economics* 68(3): 605-612.

Bringezu, S. 2011. Key elements for Economy-wide Sustainable Resource Management. *Annales des Mines - Responsabilité et environnement* 61: 78–87.

Bringezu, S. 2015. Possible Target Corridor for Sustainable Use of Global Material Resources. *Resources* 4(1): 25-54.

Bringezu, S. 2014. Targets for Global Resource Consumption. In *Factor X : policy, strategies and instruments for a sustainable resource use*, edited by M. Angrick, et al. Dordrecht: Springer.

Bringezu, S. 2009. Visions of a sustainable resource use. In *Sustainable Resource Management: Global Trends, Visions and Policies*, edited by S. Bringezu and R. Bleischwitz. Sheffield: Greenleaf Publishing Limited.

Bringezu, S., M. O'Brien, and H. Schütz. 2012. Beyond biofuels: Assessing global land use for domestic consumption of biomass. *Land Use Policy* 29(1): 224-232.

Brook, B. W., E. C. Ellis, M. P. Perring, A. W. Mackay, and L. Blomqvist. 2013. Does the terrestrial biosphere have planetary tipping points? *Trends Ecol Evol* 28(7): 396-401.

Brundtland, G., M. Khalid, S. Agnelli, S. Al-Athel, B. Chidzero, L. Fadika, V. Hauff, I. Lang, M. Shijun, and M. M. de Botero. 1987. *Report of the World Commission on Environment and Development: Our Common Future*.

Cadima, E. L. 2003. *Fish Stock Assessment Manual*. FAO Fisheries Technical Paper 393. Rome: Food and Agriculture Organization of the United Nations.

Cadotte, M. W., B. J. Cardinale, and T. H. Oakley. 2008. Evolutionary history and the effect of biodiversity on plant productivity. *Proc Natl Acad Sci U S A* 105(44): 17012-17017.

Campbell, J., J. Neuner, L. See, S. Fritz, D. Fraisl, J. Espey, and A. Kim. 2020. The role of combining national official statistics with global monitoring to close the data gaps in the environmental SDGs. *Statistical Journal of the IAOS* 36: 443-453.

Cape, J. N., J. van der Eerden, L. J. Sheppard, I. D. Leith, and M. A. Sutton. 2009. Reassessment of Critical Levels for Atmospheric Ammonia. In *Atmospheric Ammonia. Detecting emission changes and environmental impacts. Results of an Expert Workshop under the Convention on Long-range Transboundary Air Pollution*, edited by M. Sutton, et al.: Springer.

Capon, S. J., A. J. J. Lynch, N. Bond, B. C. Chessman, J. Davis, N. Davidson, M. Finlayson, P. A. Gell, D. Hohnberg, C. Humphrey, R. T. Kingsford, D. Nielsen, J. R. Thomson, K. Ward, and R. M. Nally. 2015. Regime shifts, thresholds and multiple stable states in freshwater ecosystems; a critical appraisal of the evidence. *Science of The Total Environment* 534: 122-130.

Cardinale, B. J., K. L. Matulich, D. U. Hooper, J. E. Byrnes, E. Duffy, L. Gamfeldt, P. Balvanera, M. I. O'Connor, and A. Gonzalez. 2011. The functional role of producer diversity in ecosystems. *Am J Bot* 98(3): 572-592.

Cardinale, B. J., J. E. Duffy, A. Gonzalez, D. U. Hooper, C. Perrings, P. Venail, A. Narwani, G. M. Mace, D. Tilman, D. A. Wardle, A. P. Kinzig, G. C. Daily, M. Loreau, J. B. Grace, A. Larigauderie, D. S. Srivastava, and S. Naeem. 2012. Biodiversity loss and its impact on humanity. *Nature* 486(7401): 59-67.

Carpenter, S. R., B. J. Benson, R. Biggs, J. W. Chipman, J. A. Foley, S. A. Golding, R. B. Hammer, P. C. Hanson, P. T. J. Johnson, A. M. Kamarainen, T. K. Kratz, R. C. Lathrop, K. D. McMahon, B. Provencher, J. A. Rusak, C. T. Solomon, E. H. Stanley, M. G. Turner, M. J. Vander Zanden, C.-H. Wu, and H. Yuan. 2007. Understanding Regional Change: A Comparison of Two Lake Districts. *BioScience* 57(4): 323-335.

CBD. 2010. *The Strategic Plan for Biodiversity 2011-2020 and the Aichi Biodiversity Targets* Decision X/2. Nagoya: Convention on Biological Diversity. Decision adopted by the Conference of the Parties to the Convention on Biological Diversity at its Tenth Meeting.

Ceballos, G., P. R. Ehrlich, A. D. Barnosky, A. García, R. M. Pringle, and T. M. Palmer. 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances* 1(5).

Clarke, L., K. Jiang, K. Akimoto, M. Babiker, G. Blanford, K. Fisher-Vanden, J.-C. Hourcade, V. Krey, E. Kriegler, A. Löschel, D. McCollum, S. Paltsev, S. Rose, P. R. Shukla, M. Tavoni, B. C. C. van der Zwaan, and D. P. van Vuuren. 2014. Assessing Transformation Pathways. In *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by O. Edenhofer, et al. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

CLRTAP. 2017. *Mapping critical loads for ecosystems, Chapter V of Manual on methodologies and criteria for modelling and mapping critical loads and levels and air pollution effects, risks and trends*. Berlin: UNECE Convention on Long-range Transboundary Air Pollution.

Cochrane, K. L. 2009. The Use of Scientific Information. In *A Fishery Manager's Guidebook*, edited by K. L. Cochrane and S. M. Garcia. Rome: Food and Agriculture Organization of the United Nations and Wiley-Blackwell.

Cole, M. J., R. M. Bailey, and M. G. New. 2014. Tracking sustainable development with a national barometer for South Africa using a downscaled "safe and just space" framework. *Proceedings of the National Academy of Sciences* 111(42): E4399-E4408.

Comte, A., C. Surun, and H. Levrel. 2021. *Implementation of the ESGAP framework in New-Caledonia*. Chaire Comptabilité Écologique.

Costanza, R. and H. E. Daly. 1992. Natural Capital and Sustainable Development. *Conservation Biology* 6(1): 37-46.

Daly, H. E. 1991. Elements of Environmental Macroeconomics. In *Ecological Economics: The Science and Management of Sustainability*, edited by R. Costanza. New York: Columbia University Press.

Damkjaer, S. and R. Taylor. 2017. The measurement of water scarcity: Defining a meaningful indicator. *Ambio*.

Dao, H., D. Friot, P. Peduzzi, B. Chatenoux, A. D. Bono, and S. Schwarzer. 2015. *Environmental limits and Swiss footprints based on Planetary Boundaries*. Geneva: UNEP/GRID-Geneva & University of Geneva.

Dasgupta, P. 2021. *The Economics of Biodiversity: The Dasgupta Review*. London: HM Treasury.

Dawes, J. H. P. 2020. Are the Sustainable Development Goals self-consistent and mutually achievable? *Sustainable Development* 28(1): 101-117.

De Groot, R., J. Van der Perk, A. Chiesura, and A. van Vliet. 2003. Importance and threat as determining factors for criticality of natural capital. *Ecological Economics* 44(2-3): 187-204.

de Groot, R., B. Fisher, M. Christie, J. Aronson, L. Braat, J. Gowdy, R. Haines-Young, E. Maltby, A. Neuville, S. Polasky, R. Portela, and I. Ring. 2010. Integrating the ecological and economic dimensions in biodiversity and ecosystem service valuation. In *The Economics of Ecosystems and Biodiversity: The Ecological and Economic Foundations*, edited by TEEB. London and Washington: Earthscan.

de Groot, R. S., M. A. Wilson, and R. M. J. Boumans. 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. *Ecological Economics* 41(3): 393-408.

de la Cruz, R. and J. U. Kreft. 2018. Geometric mean extension for data sets with zeros. *arXiv preprint* 1806.06403.

Desing, H., G. Braun, and R. Hischier. 2020. Ecological resource availability: a method to estimate resource budgets for a sustainable economy. *Global Sustainability* 3: e31.

Dhomse, S. S., W. Feng, S. A. Montzka, R. Hossaini, J. Keeble, J. A. Pyle, J. S. Daniel, and M. P. Chipperfield. 2019. Delay in recovery of the Antarctic ozone hole from unexpected CFC-11 emissions. *Nature Communications* 10(1): 5781.

Di Giulio, A. and D. Fuchs. 2014. Sustainable Consumption Corridors: Concept, Objections, and Responses. *GAIA - Ecological Perspectives for Science and Society* 23(3): 184-192.

Díaz, S., U. Pascual, M. Stenseke, B. Martín-López, R. T. Watson, Z. Molnár, R. Hill, K. M. A. Chan, I. A. Baste, K. A. Brauman, S. Polasky, A. Church, M. Lonsdale, A. Larigauderie, P. W. Leadley, A. P. E. van Oudenhoven, F. van der Plaat, M. Schröter, S. Lavorel, Y. Aumeeruddy-Thomas, E. Bukvareva, K. Davies, S. Demissew, G. Erpul, P. Failler, C. A. Guerra, C. L. Hewitt, H. Keune, S. Lindley, and Y. Shirayama. 2018. Assessing nature's contributions to people. *Science* 359(6373): 270-272.

Díaz, S. and M. Cabido. 2001. Vive la différence: plant functional diversity matters to ecosystem processes. *Trends in Ecology & Evolution* 16(11): 646-655.

Dickens, C., V. Smakhtin, M. McCartney, G. O'Brien, and L. Dahir. 2019. Defining and Quantifying National-Level Targets, Indicators and Benchmarks for Management of Natural Resources to Achieve the Sustainable Development Goals. 11(2): 462.

Dickens, C., M. McCartney, D. Tickner, I. J. Harrison, P. Pacheco, and B. Ndhlovu. 2020. Evaluating the Global State of Ecosystems and Natural Resources: Within and Beyond the SDGs. 12(18): 7381.

Dittrich, M., S. Giljum, S. Lutter, and C. Polzin. 2012. *Green economies around the world ? Implications of resource use for development and the environment*. Vienna.

Doherty, T. S., L. M. Bland, B. A. Bryan, T. Neale, E. Nicholson, E. G. Ritchie, and D. A. Driscoll. 2018. Expanding the Role of Targets in Conservation Policy. *Trends in Ecology & Evolution* 33(11): 809-812.

Douglass, A. and V. Fioletov. 2011. Chapter 2: Stratospheric Ozone and Surface Ultraviolet Radiation. In *Scientific Assessment of Ozone Depletion: 2010*, edited by WMO. Geneva: World Meteorological Organization,.

Drielsma, J. A., A. J. Russell-Vaccari, T. Drnek, T. Brady, P. Weihed, M. Mistry, and L. P. Simbor. 2016. Mineral resources in life cycle impact assessment—defining the path forward. *The International Journal of Life Cycle Assessment* 21(1): 85-105.

DW. 2020. Climate activists protest Germany's new Datteln 4 coal power plant. *Deutsche Welle*, section.

EC. 2002. *Proposal for a Directive of the European Parliament and of the Council concerning the quality of bathing water*. COM(2002) 581 final. Brussels: European Commission.

EC. 2003. *Overall approach to the classification of ecological status and ecological potential*. No. 13. Brussels: European Commission. Common Implementation Strategy for the Water Framework Directive (2000/60/EC) Guidance Report.

EC. 2005. *Guidance on the intercalibration process 2004 - 2006*. No. 14. Brussels: European Commission. Common Implementation Strategy for the Water Framework Directive (2000/60/EC) Guidance Report.

EC. 2006. *Thematic Strategy for Soil Protection*. COM(2006) 231. Brussels: European Commission.

EC. 2008. *Addressing the challenges of deforestation and forest degradation to tackle climate change and biodiversity loss*. COM(2008) 645. Brussels: European Commission.

EC. 2009. *Guidance Document No. 18. Guidance on Grounwater Status and Trend Assessment*. Technical Report - 2009 - 026. Luxembourg: European Commission.

EC. 2010. Commission Decision of 1 September 2010 on criteria and methodological standards on good environmental status of marine waters. Brussels.

EC. 2011a. *Guidance Document No. 27 - Technical Guidance For Deriving Environmental Quality Standards*. Brussels: European Commission.

EC. 2011b. *Links between the Water Framework Directive and Nature Directives*. Brussels: European Commission.

EC. 2013. *The Economic benefits of the Natura 2000 Network*. Brussels: European Commission.

EC. 2014a. *The first phase of implementation of the Marine Strategy Framework Directive (2008/56/EC) - The European Commission's assessment and guidance*. SWD(2014) 49 final. Brussels: European Commission.

EC. 2014b. *General Union Environment Action Programme to 2020: Living well, within the limits of our planet*. Brussels: European Commission.

EC. 2016. *Synthesis Report on the Quality of Drinking Water in the Union examining Member States' reports for the 2011-2013 period, foreseen under Article 13(5) of Directive 98/83/EC*. COM(2016) 666 final. Brussels.

EC. 2017. Commission Decision (EU) 2017/848 of 17 May 2017 laying down criteria and methodological standards on good environmental status of marine waters and specifications and standardised methods for monitoring and assessment, and repealing Decision 2010/477/EU Brussels.

EC. 2018. *Proposal for a Directive of the European Parliament and of the Council on the quality of water intended for human consumption (recast)*. COM(2017) 753 final. Brussels: European Commission.

EC. 2019a. *Natural and Cultural Heritage in Europe: Working together within the Natura 2000 network*. Brussels: European Commission.

EC. 2019b. *The European Green Deal*. COM(2019) 640 final. Brussels: European Commission.

EC. 2020. *Key stages and progress up to 2019 - Accompanying the Report from the Commission to the European Parliament and the Council on the implementation of the Marine Strategy Framework Directive (Directive 2008/56/EC)*. SWD(2020) 60 final. Brussels: European Commission.

ECA. 2018. *Air pollution: Our health still insufficiently protected*. 23. Luxembourg: European Court of Auditors.

Eckelmann, W., R. Baritz, S. Bialousz, P. Bielek, F. Carré, B. Houšková, R. J. A. Jones, M. Kibblewhite, J. Kozak, C. L. Bas, G. Tóth, T. Tóth, G. Várallyay, Markku Yli Halla, and M. Zupan. 2006. *Common Criteria for Risk Area Identification according to Soil Threats*. European Soil Bureau Research Report No.20. Luxembourg: Joint Research Centre - Institute for Environment and Sustainability.

Eckert, E. and O. Kovalevska. 2021. Sustainability in the European Union: Analyzing the Discourse of the European Green Deal. *Journal of Risk and Financial Management* 14(2).

ECOSOC. 2018. *Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators*. E/CN.3/2018/2. United Nations Economic and Social Council.

EEA. 1999. *Environmental indicators: Typology and overview*. No 25. Copenhagen: European Environment Agency. Technical report.

EEA. 2003a. *Indicator Fact Sheet: (WQ01c) Water exploitation index*. Copenhagen: European Environment Agency.

EEA. 2003b. *Environmental Indicators: Typology and Use in Reporting*. Copenhagen: European Environment Agency. EEA internal working paper.

EEA. 2010. *Forest: growing stock, increment and fellings: Indicator Assessment*. Copenhagen: European Environment Agency.

EEA. 2015. *The European Environment - State and Outlook 2015: Synthesis*. Copenhagen: European Environment Agency.

EEA. 2016a. *Air quality in Europe — 2016 report*. No 29/2016. Copenhagen: European Environment Agency. EEA Report.

EEA. 2016b. *Use of freshwater resources: Indicator assessment*. Copenhagen: European Environment Agency.

EEA. 2017. *Forest: growing stock, increment and fellings*. Copenhagen: European Environment Agency.

EEA. 2018a. *Status of the assessed European fish stocks in relation to Good Environmental Status per regional sea*. Copenhagen: European Environment Agency.

EEA. 2018b. *European waters. Assessment of status and pressures 2018*. No 7/2018. Copenhagen: European Environment Agency. EEA Report.

EEA. 2018c. *Use of freshwater resources*. Copenhagen: European Environment Agency.

EEA. 2019a. *Production and consumption of ozone-depleting substances in Europe*. Copenhagen: European Environment Agency.

EEA. 2019b. *Ozone-depleting substances 2019. Aggregated data reported by companies on the import, export, production, destruction, feedstock and process agent use of ozone-depleting substances in the European Union, 2006-2018*. No 12/2019. Copenhagen: European Environment Agency. EEA Report.

EEA. 2019c. *The European environment — state and outlook 2020. Knowledge for transition to a sustainable Europe*. Copenhagen: European Environment Agency.

EEA. 2019d. *Status of the assessed European commercial fish and shellfish stocks in relation to Good Environmental Status (GES) per EU marine region in 2015-2017*. Copenhagen: European Environment Agency.

EEA. 2019e. *Air quality in Europe — 2019 report*. No 10/2019. Copenhagen: European Environment Agency. EEA Report.

EEA. 2019f. *Country reports 2018 bathing season*. Copenhagen: European Environment Agency.
EEA. 2020a. *National summary dashboards - Habitats Directive – Art.17*. Copenhagen: European Environment Agency.
EEA. 2020b. *State of nature in the EU. Results from reporting under the nature directives 2013-2018*. EEA Report No 10/2020. Copenhagen: European Environment Agency.
EEA. 2020c. *European bathing water quality in 2019*. Briefing No 1/2020. Copenhagen: European Environment Agency.
EEA. 2020d. *Nutrients in freshwater in Europe*. Copenhagen: European Environment Agency.
Eisenmenger, N., M. Pichler, N. Krenmayr, D. Noll, B. Plank, E. Schalmann, M.-T. Wandl, and S. Gingrich. 2020. The Sustainable Development Goals prioritize economic growth over sustainable resource use: a critical reflection on the SDGs from a socio-ecological perspective. *Sustainability Science* 15(4): 1101-1110.
Ekins, P. 2000. *Economic Growth and Environmental Sustainability: the Prospects for Green Growth*. London/New York: Routledge.
Ekins, P. 2001. From green GNP to the sustainability gap: Recent developments in national environmental economic accounting. *Journal of Environmental Assessment Policy and Management* 03(01): 61-93.
Ekins, P. 2011. Environmental sustainability: From environmental valuation to the sustainability gap. *Progress in Physical Geography* 35(5): 629-651.
Ekins, P. 1992. A Four-Capital Model of Wealth Creation. In *Real-Life Economics: Understanding Wealth-Creation*, edited by P. Ekins and M. Max-Neef. London: Routledge.
Ekins, P. and S. Simon. 1999. The sustainability gap: a practical indicator of sustainability in the framework of the national accounts. *International Journal of Sustainable Development* 2(1): 32-58.
Ekins, P. and S. Simon. 2001. Estimating sustainability gaps: methods and preliminary applications for the UK and the Netherlands. *Ecological Economics* 37(1): 5-22.
Ekins, P. and S. Simon. 2003. An illustrative application of the CRITINC framework to the UK. *Ecological Economics* 44(2-3): 255-275.
Ekins, P. and A. Usubiaga. 2019. Brundtland+30: the continuing need for an indicator of environmental sustainability. In *What Next for Sustainable Development? Our Common Future at Thirty*, edited by J. Meadowcroft, et al.: Edward Elgar Publishing.
Ekins, P. and S. Simon. 1998. Determining the Sustainability Gap: National Accounting for Environmental Sustainability. In *UK Environmental Accounts: Theory, Data and Application*, edited by P. Vaze. London: Office for National Statistics.
Ekins, P., C. Folke, and R. De Groot. 2003a. Identifying critical natural capital. *Ecological Economics* 44(2-3): 159-163.
Ekins, P., B. Meyer, and F. Schmidt-Bleek. 2009. *Reducing Resource Consumption. A Proposal for Global Resource and Environmental Policy*. gws Discussion Paper 2009/5. Osnabrück: Gesellschaft für Wirtschaftliche Strukturforschung GmbH.
Ekins, P., B. Milligan, and A. Usubiaga-Liaño. 2019. *A single indicator of strong sustianability for development: Theoretical basis and practical implementation*. No. 2019-122. Agence Française de Développement. AFD Research Papers.
Ekins, P., S. Simon, L. Deutsch, C. Folke, and R. De Groot. 2003b. A framework for the practical application of the concepts of critical natural capital and strong sustainability. *Ecological Economics* 44(2-3): 165-185.
Ekvall, T., M. Elander, K. Umpfenbach, M. Hirschnitz-Garbers, C. Hudson, S. Wunder, M. Nesbit, C. Keenleyside, L. Mazza, D. Russi, G. Tucker, E. Underwood, S. Withana, M. Bicket, R. Vanner, M. A. Kong, A. Tan, A. Bigano, F. Eboli, J. Gaska, and A. Sniegocki. 2015. *Development of DYNAMIX Policy Mixes*. D4.2. DYNAMIX project.
Elder, M. and S. H. Olsen. 2019. The Design of Environmental Priorities in the SDGs. *Global Policy* 10(S1): 70-82.

Ellis, E. C. 2018. *Anthropocene: A Very Short Introduction*: Oxford University Press,.

Erb, K. H., H. Haberl, R. DeFries, E. C. Ellis, F. Krausmann, and P. H. Verburg. 2012. Pushing the planetary boundaries. *Science* 338(6113): 1419-1420; author reply 1420.

Esty, D. C. 2018. Measurement Matters: Toward Data-Driven Environmental Policy-making. In *Routledge Handbook of Sustainability Indicators*, edited by Simon Bell and Stephen Morse. New York: Routledge.

European Council. 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora.

European Council. 1998. Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption.

European Parliament and European Council. 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy.

European Parliament and European Council. 2006. Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC.

European Parliament and European Council. 2008a. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive).

European Parliament and European Council. 2008b. Directive 2008/105/EC on environmental quality standards in the field of water policy, amending and subsequently repealing Council Directives 82/176/EEC, 83/513/EEC, 84/156/EEC, 84/491/EEC, 86/280/EEC and amending Directive 2000/60/EC of the European Parliament and of the Council.

European Parliament and Council of the European Union. 2009. *Regulation (EC) No 1005/2009 of the European Parliament and of the Council of 16 September 2009on substances that deplete the ozone layer*. Brussels.

European Parliament and Council of the European Union. 2013. Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy. Brussels.

Eurostat. 2014a. *Getting messages across using indicators. A handbook based on experiences from assessing Sustainable Development Indicators*. Luxembourg: Eurostat.

Eurostat. 2014b. *Towards a harmonised methodology for statistical indicators. Part 1: Indicator typologies and terminologies*. Luxembourg: Eurostat.

Eurostat. 2015. Water exploitation index (tsdnr310). Indicator Profile (ESMS).

Eurostat. 2019a. *Greenhouse gas emissions by source sector*. Luxembourg: Eurostat.

Eurostat. 2019b. *Sustainable development in the European Union. Monitoring report on progress towards the SDGs in an EU context*. Luxembourg: Eurostat.

Eurostat. 2020a. *EU SDG Indicator set 2020. Result of the review in preparation of the 2020edition of the EU SDG monitoring report*. Luxembourg: Eurostat.

Eurostat. 2020b. *Sustainable development in the European Union. Monitoring report on progress towards the SDGs in an EU context*. Luxembourg: Eurostat.

Eurostat. 2021. Renewable freshwater resources (env_wat_res).

Faergemann, H. 2012. *Update on Water Scarcity and Droughts indicator development*. DG Environment.

Fagerli, H., S. Tsyro, J. E. Jonson, Á. Nyíri, D. Simpson, P. Wind, A. Benedictow, H. Klein, Q. Mu, B. R. Denby, E. G. Wærsted, W. Aas, S. Eckhardt, A. Hjellbrekke, S. Solberg, S. M. Platt, K. E. Yttri, K. Tørseth, K. Mareckova, B. Matthews, S. Schindlbacher, B. Ullrich, R. Wankmüller, T. Scheuschner, R. Bergström, H. A.C., D. v. d. Gon, J. J. P. Kuenen, A. J. H. Visschedijk, S. Reimann, M. Hill, and A. Claude. 2020. *Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components. Status Report 2020*. EMEP Report 1/2020. Oslo: Norwegian Meteorological Institute.

Fairbrass, A., G. Mace, P. Ekins, and B. Milligan. 2020a. The natural capital indicator framework (NCIF) for improved national natural capital reporting. *Ecosystem Services* 46: 101198.

Fairbrass, A., A. Usubiaga-Liaño, P. Ekins, and B. Milligan. 2020b. *Data opportunities and challenges for calculating a global Strong Environmental Sustainability (SES) index*. No. 133. Agence Française de Développement. AFD Research Papers.

Falkenmark, M. 1989. The massive water scarcity now threatening Africa: why isn't it being addressed? *Ambio*: 112-118.

Falkenmark, M. and G. Lindh. 1976. *Water for a Starving World*: Westview Press.

Falkenmark, M., J. Lundqvist, and C. Widstrand. 1989. Macro-scale water scarcity requires micro-scale approaches. Aspects of vulnerability in semi-arid development. *Nat Resour Forum* 13(4): 258-267.

FAO. 2011. *Review of the state of world marine fishery resources*. 569. Rome: Food and Agriculture Organization of the United Nations. FAO Fisheries and Aquaculture Technical Paper.

FAO. 2015. *Status of the World's Soil Resources - Main Report*. Rome: Food and Agriculture Organization of the United Nations.

FAOSTAT. 2020. *Forestry Production and Trade*. Food and Agriculture Organization of the United Nations.

Flynn, D. F. B., N. Mirotchnick, M. Jain, M. I. Palmer, and S. Naeem. 2011. Functional and phylogenetic diversity as predictors of biodiversity–ecosystem-function relationships. *Ecology* 92: 1573-1581.

Forest Europe. 2020. *State of Europe's Forests 2020*. Bratislava: Ministerial Conference on the Protection of Forests in Europe.

Forest Europe, UNECE, and FAO. 2007. *State of Europe's Forests 2007. The MCPFE report on sustainable forest management in Europe*. Warsaw: Ministerial Conference on the Protection of Forests in Europe.

Forest Europe, UNECE, and FAO. 2011. *State of Europe's Forests 2011. Status and Trends in Sustainable Forest Management in Europe*. Oslo: Ministerial Conference on the Protection of Forests in Europe.

Forest Europe, UNECE, and FAO. 2015. *State of Europe's Forests 2015*. Madrid: Ministerial Conference on the Protection of Forests in Europe.

Friedlingstein, P., R. M. Andrew, J. Rogelj, G. P. Peters, J. G. Canadell, R. Knutti, G. Luderer, M. R. Raupach, M. Schaeffer, D. P. van Vuuren, and C. Le Quéré. 2014. Persistent growth of CO2 emissions and implications for reaching climate targets. *Nature Geoscience* 7(10): 709-715.

Fuhrer, J., L. Skärby, and M. R. Ashmore. 1997. Critical levels for ozone effects on vegetation in Europe. *Environmental Pollution* 97(1): 91-106.

Fuhrer, J., M. R. Ashmore, G. Mills, F. Hayes, and A. W. Davison. 2003. Critical levels for semi-natural vegetation. In *Establishing Ozone Critical Levels II, UNECE Workshop Report*, edited by P.-E. Karlsson, et al. Gothenburg.

Gagic, V., I. Bartomeus, T. Jonsson, A. Taylor, C. Winqvist, C. Fischer, E. M. Slade, I. Steffan-Dewenter, M. Emmerson, S. G. Potts, T. Tscharntke, W. Weisser, and R. Bommarco. 2015. Functional identity and diversity of animals predict ecosystem functioning better than species-based indices. *Proceedings of the Royal Society B: Biological Sciences* 282(1801): 20142620.

Geary, R. S., B. Wheeler, R. Lovell, R. Jepson, R. Hunter, and S. Rodgers. 2021. A call to action: Improving urban green spaces to reduce health inequalities exacerbated by COVID-19. *Preventive Medicine* 145: 106425.

Gejl, R. N., P. L. Bjerg, H. J. Henriksen, M. Z. Hauschild, J. Rasmussen, and M. Rygaard. 2018. Integrating groundwater stress in life-cycle assessments – An evaluation of water abstraction. *Journal of Environmental Management* 222: 112-121.

Gerlagh, R., R. Dellink, M. Hofkes, and H. Verbruggen. 2002. A measure of sustainable national income for the Netherlands. *Ecological Economics* 41(1): 157-174.

Gerten, D., H. Hoff, J. Rockström, J. Jägermeyr, M. Kummu, and A. V. Pastor. 2013. Towards a revised planetary boundary for consumptive freshwater use: role of environmental flow requirements. *Current Opinion in Environmental Sustainability* 5(6): 551-558.

Giampietro, M. and A. Saltelli. 2014a. Footworking in circles: Reply to Goldfinger et al. (2014) "Footprint Facts and Fallacies: A Response to Giampietro and Saltelli (2014) Footprints to nowhere". *Ecological Indicators* 46: 260-263.

Giampietro, M. and A. Saltelli. 2014b. Footprints to nowhere. *Ecological Indicators* 46: 610-621.

Giannetti, B. F., F. Agostinho, C. M. V. B. Almeida, and D. Huisingh. 2015. A review of limitations of GDP and alternative indices to monitor human wellbeing and to manage eco-system functionality. *Journal of Cleaner Production* 87: 11-25.

Gleeson, T., L. Wang-Erlandsson, S. C. Zipper, M. Porkka, F. Jaramillo, D. Gerten, I. Fetzer, S. E. Cornell, L. Piemontese, L. J. Gordon, J. Rockström, T. Oki, M. Sivapalan, Y. Wada, K. A. Brauman, M. Flörke, M. F. P. Bierkens, B. Lehner, P. Keys, M. Kummu, T. Wagener, S. Dadson, T. J. Troy, W. Steffen, M. Falkenmark, and J. S. Famiglietti. 2020. The Water Planetary Boundary: Interrogation and Revision. *One Earth* 2(3): 223-234.

Gleeson, T., L. W. Erlandsson, S. C. Zipper, M. Porkka, F. Jaramillo, D. Gerten, I. Fetzer, S. E. Cornell, L. Piemontese, L. Gordon, J. Rockström, T. Oki, M. Sivapalan, Y. Wada, K. A. Brauman, M. Flörke, M. F. P. Bierkens, B. Lehner, P. Keys, M. Kummu, T. Wagener, S. Dadson, T. J. Troy, W. Steffen, M. Falkenmark, and J. S. Famiglietti. submitted. The water planetary boundary: a roadmap to illuminate water cycle modifications in the Anthropocene. *Water Resources Research*.

González-Fernández, I., E. Calvo, G. Gerosa, V. Bermejo, R. Marzuoli, V. Calatayud, and R. Alonso. 2014. Setting ozone critical levels for protecting horticultural Mediterranean crops: Case study of tomato. *Environmental Pollution* 185: 178-187.

Goodland, R. 1995. The Concept of Environmental Sustainability. *Annual Review of Ecology and Systematics* 26: 1-24.

Greco, S., A. Ishizaka, M. Tasiou, and G. Torrisi. 2019. On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. *Social Indicators Research* 141(1): 61-94.

Greenland, D. J., D. Rimmer, and D. Payne. 1975. DETERMINATION OF THE STRUCTURAL STABILITY CLASS OF ENGLISH AND WELSH SOILS, USING A WATER COHERENCE TEST. *Journal of Soil Science* 26(3): 294-303.

Gregorich, E. G., C. M. Monreal, M. R. Carter, D. A. Angers, and B. H. Ellert. 1994. Towards a minimum data set to assess soil organic matter quality in agricultural soils. *Canadian Journal of Soil Science* 74(4): 367-385.

Griggs, D., M. Stafford-Smith, O. Gaffney, J. Rockstrom, M. C. Ohman, P. Shyamsundar, W. Steffen, G. Glaser, N. Kanie, and I. Noble. 2013. Policy: Sustainable development goals for people and planet. *Nature* 495(7441): 305-307.

Guinée, J. B. and R. Heijungs. 1995. A proposal for the definition of resource equivalency factors for use in product life-cycle assessment. *Environmental Toxicology and Chemistry* 14(5): 917-925.

Haberl, H., K.-H. Erb, and F. Krausmann. 2014. Human Appropriation of Net Primary Production: Patterns, Trends, and Planetary Boundaries. *Annual Review of Environment and Resources* 39(1): 363-391.

Hák, T., S. Janoušková, and B. Moldan. 2016. Sustainable Development Goals: A need for relevant indicators. *Ecological Indicators* 60: 565-573.

Hametner, M. and M. Kostetckaia. 2020. Frontrunners and laggards: How fast are the EU member states progressing towards the sustainable development goals? *Ecological Economics* 177: 106775.

Hansen, J. 2005. A slippery slope: How much global warming constitutes "dangerous anthropogenic interference"? *Climatic Change* 68(3): 269-279.

Hansen, J., M. Sato, P. Kharecha, D. Beerling, R. Berner, V. Masson-Delmotte, M. Pagani, M. Raymo, D. L. Royer, and J. C. Zachos. 2008. Target atmospheric CO2: Where should humanity aim? *The Open Atmospheric Science Journal* 2(1): 217--231.

Häyhä, T., P. L. Lucas, D. P. van Vuuren, S. E. Cornell, and H. Hoff. 2016. From Planetary Boundaries to national fair shares of the global safe operating space — How can the scales be bridged? *Global Environmental Change* 40: 60-72.

Heck, V., D. Gerten, W. Lucht, and A. Popp. 2018a. Biomass-based negative emissions difficult to reconcile with planetary boundaries. *Nature Climate Change* 8(2): 151-155.

Heck, V., H. Hoff, S. Wirsenius, C. Meyer, and H. Kreft. 2018b. Land use options for staying within the Planetary Boundaries – Synergies and trade-offs between global and local sustainability goals. *Global Environmental Change* 49: 73-84.

Helsel, D. R. 2005. More Than Obvious: Better Methods for Interpreting Nondetect Data. *Environmental Science & Technology* 39(20): 419A-423A.

Henckens, M. L. C. M., P. P. J. Driessen, and E. Worrell. 2014. Metal scarcity and sustainability, analyzing the necessity to reduce the extraction of scarce metals. *Resources, Conservation and Recycling* 93: 1-8.

Henriksen, H. J., L. Troldborg, A. L. Højberg, and J. C. Refsgaard. 2008. Assessment of exploitable groundwater resources of Denmark by use of ensemble resource indicators and a numerical groundwater–surface water model. *Journal of Hydrology* 348(1): 224-240.

Henriksen, H. J., J. Rasmussen, M. Olsen, X. He, L. Troldborg, and L. F. Jørgensen. 2014. *Implementering Af Modeller Til Brug for Vandforvaltning . Delprojekt: Effekt Af Vandindvinding Konceptuel Tilgang Og Validering Samt Tilstandsvurdering Af Grundvandsforekomster Udarbejdet for Naturstyrelsen*.

Hettelingh, J.-P., M. Posch, and J. Slootweg. 2017. *European critical loads: database, biodiversity and ecosystems at risk. CCE Final Report 2017*. RIVM Report 2017-0155. Bilthoven: Coordination Centre for Effects.

Hettelingh, J.-P., G. Schütze, W. de Vries, H. Denier van der Gon, I. Ilyin, G. J. Reinds, J. Slootweg, and O. Travnikov. 2015. Critical Loads of Cadmium, Lead and Mercury and Their Exceedances in Europe. In *Critical Loads and Dynamic Risk Assessments: Nitrogen, Acidity and Metals in Terrestrial and Aquatic Ecosystems*, edited by W. de Vries, et al. Dordrecht: Springer Netherlands.

Hoekstra, A. Y. and M. M. Mekonnen. 2011. *Global water scarcity: The monthly blue water footprint compared to blue water availability for the world's major river basins*. 53. Delft: UNESCO-IHE Institute for Water Education. Value of Water Research Report Series.

Hoekstra, A. Y., A. K. Chapagain, M. M. Aldaya, and M. M. Mekonnen. 2011. *The water footprint assessment manual: Setting the global standard*. London, Washington DC: Routledge.

Hoff, H., B. Nykvist, and M. Carson. 2014. *"Living well, within the limits of our planet"? Measuring Europe's growing external footprint*. Working Paper 2014-05. Stockholm: Stockholm Environment Institute.

Höhne, N., M. den Elzen, and D. Escalante. 2014. Regional GHG reduction targets based on effort sharing: a comparison of studies. *Climate Policy* 14(1): 122-147.

Holdren, J. P., G. C. Daily, and P. R. Ehrlich. 1995. The meaning of sustainability: biogeophysical aspects. In *Defining and Measuring Sustainability*, edited by M. Munasingha and W. Shearer. Washington, D.C.: World Bank,.

Holland, M. R., P. W. Mueller, A. J. Rutter, and P. J. A. Shaw. 1995. Growth of coniferous trees exposed to SO2 and O3 using an open-air fumigation system. *Plant, Cell & Environment* 18(3): 227-236.

Holt, S. J. and L. M. Talbot. 1978. New Principles for the Conservation of Wild Living Resources. *Wildlife Monographs*(59): 3-33.

Hooper, D. U., E. C. Adair, B. J. Cardinale, J. E. K. Byrnes, B. A. Hungate, K. L. Matulich, A. Gonzalez, J. E. Duffy, L. Gamfeldt, and M. I. O/'Connor. 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* 486(7401): 105-108.

Horálek, J., M. Schreiberová, F. d. Leeuw, P. Kurfürst, P. d. Smet, and J. Schovánková. 2019. *European air quality maps for 2016: PM10, PM2.5, Ozone, NO2 and NOx spatial estimates and their uncertainties*. ETC/ACM 2018/8. Bilthoven: European Topic Centre on Air Pollution and Climate Change Mitigation.

Horálek, J., M. Schreiberová, P. Schneider, P. Kurfürst, J. Schovánková, and J. Ďoubalová. 2020. *European air quality maps for 2017: PM10, PM2.5, Ozone, NO2 and NOx spatial estimates and their uncertainties*. ETC/ATNI 2019/9. Kjeller: t. European Topic Centre on Air pollution, noise and industrial pollution,.

Hsu, A., L. A. Johnson, and A. Lloyd. 2013. *Measuring Progress: A Practical Guide From the Developers of the Environmental Performance Index (EPI)*. New Haven: Yale Center for Environmental Law & Policy.

Huan, Y., H. Li, and T. Liang. 2019. A New Method for the Quantitative Assessment of Sustainable Development Goals (SDGs) and a Case Study on Central Asia. Sustainability. *Sustainability* 11: 3504.

Huber, S., G. Prokop, D. Arrouays, G. Banko, A. Bispo, R. J. A. Jones, M. G. Kibblewhite, W. Lexer, A. Möller, R. J. Rickson, T. Shishkov, M. Stephens, G. Toth, J. J. H. Van den Akker, G. Varallyay, F. G. A. Verheijen, and A. R. Jones. 2008. *Environmental Assessment of Soil for Monitoring. Volume I: Indicators & Criteria*. Luxembourg: Joint Research Centre - Institute for Environment and Sustainability.

Hueting, R. 1980. *New Scarcity and Economic Growth. More Welfare Through Less Production?* Amsterdam: North-Holland.

Hueting, R. and B. De Boer. 2001. Environmental valuation and sustainable national income according to Hueting. In *Economic growth and valuation of the environment: a debate*, edited by E. C. van Ierland, et al. Cheltenham: Edward Elgar.

Hughes, T. P., S. Carpenter, J. Rockstrom, M. Scheffer, and B. Walker. 2013. Multiscale regime shifts and planetary boundaries. *Trends Ecol Evol* 28(7): 389-395.

IAEG-SDGs. 2016. *Compilation of Metadata for the Proposed Global Indicators for the Review of the 2030 Agenda for Sustainable Development - Goal 15 Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss*. I.-a. E. G. o. S. Indicators.

IAEG-SDGs. 2018. *Indicator 15.3.1: Proportion of land that is degraded over total land area - Metadata*. Inter-agency Expert Group on SDG Indicators.

IAEG-SDGs. 2019. IAEG-SDGs Tier Classification for Global SDG Indicators. https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/. Accessed.

IAEG-SDGs. 2021. *Indicator 6.4.2: Level of water stress: freshwater withdrawal as a proportion of available freshwater resources - Metadata*. Inter-agency Expert Group on SDG Indicators.

ICES. 2003. *Report of the Study Group on Precautionary Reference Points For Advice on Fishery Management*. Copenhagen: International Council for the Exploration of the Sea.

ICES. 2018. *ICES Advice basis*. Copenhagen,: International Council for the Exploration of the Sea.

ICSU and ISSC. 2015. *Review of the Sustainable Development Goals: The Science Perspective*. Paris: International Council for Science.

IPCC. 2018. *Global Warming of 1.5°C.An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. Edited by V. Masson-Delmotte, et al. Geneva: Intergovernmental Panel on Climate Change,.

IPCC. 2013. Summary for Policymakers. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker, et al. Cambridge, United Kingdom and New York, NY, USA.

IRP. 2014. *Managing and conserving the natural resource base for sustained economic and social development*. International Resource Panel.

IRP. 2019. *Global Resources Outlook 2019: Natural Resources for the Future We Want*. Nairobi: International Resource Panel.

Jäger, J. 2014. *A Vision for a Resource Efficient Economy*. D2.2. POLFREE project.

Jalkanen, J., H. Fabritius, K. Vierikko, A. Moilanen, and T. Toivonen. 2020. Analyzing fair access to urban green areas using multimodal accessibility measures and spatial prioritization. *Applied Geography* 124: 102320.

Janoušková, S., T. Hák, and B. Moldan. 2018. Global SDGs Assessments: Helping or Confusing Indicators? *Sustainability* 10(5): 1540.

Janzen, H. H., F. J. Larney, and B. M. Olson. 1992. *Soil quality factors of problem soils in Alberta*. Vol. 29. Alberta. Proceedings of the Alberta Soil Science Workshop.

Johnson, L. C. 1987. Soil loss tolerance: Fact or myth? *Journal of Soil and Water Conservation* 42(3): 155-160.

Jones, K. R., O. Venter, R. A. Fuller, J. R. Allan, S. L. Maxwell, P. J. Negret, and J. E. Watson. 2018. One-third of global protected land is under intense human pressure. *Science* 360(6390): 788-791.

Jones, R. J. A., Y. L. Bissonnais, P. Bazzoffi, J. Sánchez, O. D. Díaz, G. Loj, L. Øygarden, V. Prasuhn, B. Rydell, P. Strauss, J. B. Üveges, L. Vandekerckhove, and Y. Yordanov. 2004. Soil Erosion. In *Reports of the Technical Working Groups Established under the Thematic Strategy for Soil Protection. Volume II: Erosion*, edited by L. Van-Camp, et al. Luxembourg: Joint Research Centre - Institute for Environment and Sustainability,.

JRC-IES. 2011. *International Reference Life Cycle Data System (ILCD) Handbook- Recommendations for Life Cycle Impact Assessment in the European context*. Ispra: Joint Research Centre - Institute for Environment and Sustainability.

JRC. 2019. *Your 10-Step Pocket Guideto Composite Indicators & Scoreboards*. Ispra: Joint Research Centre.

Kabisch, N., M. Strohbach, D. Haase, and J. Kronenberg. 2016. Urban green space availability in European cities. *Ecological Indicators* 70: 586-596.

Kågeson, P. 2000. Is Factor 10 a useful Tool in Environmental Policy? In *Dematerialisation and Factor 10*, edited by S. E. P. Agency. Stockholm: Swedish Environmental Protection Agency.

Kalf, F. R. P. and D. R. Woolley. 2005. Applicability and methodology of determining sustainable yield in groundwater systems. *Hydrogeology Journal* 13(1): 295-312.

Karlsson, P. E., J. Uddling, S. Braun, M. Broadmeadow, S. Elvira, B. S. Gimeno, D. Le Thiec, E. Oksanen, K. Vandermeiren, M. Wilkinson, and L. Emberson. 2003. New Critical Levels for Ozone Impact on Trees Based on AOT40 and Leaf Cumulated Ozone Uptake. In *Establishing Ozone Critical Levels II, UNECE Workshop Report*, edited by P.-E. Karlsson, et al. Gothenburg.

Karlsson, P. E., S. Braun, M. Broadmeadow, S. Elvira, L. Emberson, B. S. Gimeno, D. Le Thiec, K. Novak, E. Oksanen, M. Schaub, J. Uddling, and M. Wilkinson. 2007. Risk assessments for forest trees: The performance of the ozone flux versus the AOT concepts. *Environmental Pollution* 146(3): 608-616.

Kay, B. D. and D. A. Angers. 1999. Soil structure. In *Handbook of Soil Science*, edited by M. E. Sumner: CRC Press.

Kay, D. and J. Fawell. 2007. *An FWR Guide to Standards for Recreational Water Quality*. Marlow (UK): Foundation for Water Research.

Kemper, W. D. and E. J. Koch. 1966. *Aggregate stability of soils from Western United States and Canada : measurement procedure, correlations with soil constituents*. Washington, D.C.: United States Department of Agriculture.

Kenny, D. C., R. Costanza, T. Dowsley, N. Jackson, J. Josol, I. Kubiszewski, H. Narulla, S. Sese, A. Sutanto, and J. Thompson. 2019. Australia's Genuine Progress Indicator Revisited (1962–2013). *Ecological Economics* 158: 1-10.

Klinglmair, M., S. Sala, and M. Brandão. 2014. Assessing resource depletion in LCA: a review of methods and methodological issues. *The International Journal of Life Cycle Assessment* 19(3): 580-592.

Knutti, R., J. Rogelj, J. Sedlacek, and E. M. Fischer. 2016. A scientific critique of the two-degree climate change target. *Nature Geosci* 9(1): 13-18.

Körschens, M., A. Weigel, and E. Schulz. 1998. Turnover of soil organic matter (SOM) and long-term balances — tools for evaluating sustainable productivity of soils. *Zeitschrift für Pflanzenernährung und Bodenkunde* 161(4): 409-424.

Krull, E. S., J. O. Skjemstad, and J. A. Baldock. 2004. *Functions of Soil Organic Matter and the Effect on Soil Properties*. Kingston: Grains Research and Development Corporation.

Kubiszewski, I., R. Costanza, C. Franco, P. Lawn, J. Talberth, T. Jackson, and C. Aylmer. 2013. Beyond GDP: Measuring and achieving global genuine progress. *Ecological Economics* 93: 57-68.

Kutlar Joss, M., M. Eeftens, E. Gintowt, R. Kappeler, and N. Künzli. 2017. Time to harmonize national ambient air quality standards. *International Journal of Public Health* 62(4): 453-462.

Kwatra, S., A. Kumar, and P. Sharma. 2020. A critical review of studies related to construction and computation of Sustainable Development Indices. *Ecological Indicators* 112: 106061.

Lafortune, G. and G. Schmidt-Traub. 2019. *Exposing EU policy gaps to address the Sustainable Development Goals*. Brussels: European Economic and Social Committee.

Lafortune, G., G. Fuller, G. Schmidt-Traub, and C. Kroll. 2020. How Is Progress towards the Sustainable Development Goals Measured? Comparing Four Approaches for the EU. 12(18): 7675.

Lafortune, G., G. Fuller, J. Moreno, G. Schmidt-Traub, and C. Kroll. 2018. *SDG Index and Dashboards. Detailed Methodological paper*. New York: Bertelsmann Stiftung and the Sustainable Development Solutions Network (SDSN).

Lange, G.-M., Q. Wodon, and K. Carey. 2018. *The Changing Wealth of Nations 2018: Building a Sustainable Future*. Washington, DC: World Bank.

Larkin, P. A. 1977. An Epitaph for the Concept of Maximum Sustained Yield. *Transactions of the American Fisheries Society* 106(1): 1-11.

Lawn, P. A. 2003. A theoretical foundation to support the Index of Sustainable Economic Welfare (ISEW), Genuine Progress Indicator (GPI), and other related indexes. *Ecological Economics* 44(1): 105-118.

Lawrence, D. and K. Vandecar. 2014. Effects of tropical deforestation on climate and agriculture. *Nature Climate Change* 5: 27.

Li, L., S. Du, L. Wu, and G. Liu. 2009. An overview of soil loss tolerance. *CATENA* 78(2): 93-99.

Lin, D., L. Hanscom, J. Martindill, M. Borucke, L. Cohen, A. Galli, E. Lazarus, G. Zokai, K. Iha, D. Eaton, and M. Wackernagel. 2016. *Working Guidebook to the National Footprint Accounts: 2016 Edition*. Oakland: Global Footprint Network.

Logan, T. J. 1982. Improved Criteria for Developing Soil Loss Tolerance Levels for Cropland. In *Determinants of Soil Loss Tolerance*, edited by American Society of Agronomy and Soil Science Society of Ametica. Madison.

Louette, G., D. Adriaens, D. Paelinckx, and M. Hoffmann. 2015. Implementing the Habitats Directive: How science can support decision making. *Journal for Nature Conservation* 23: 27-34.

Lovejoy, T. E. and C. Nobre. 2018. Amazon Tipping Point. *Science Advances* 4(2): eaat2340.

Loveland, P. and J. Webb. 2003. Is there a critical level of organic matter in the agricultural soils of temperate regions: a review. *Soil and Tillage Research* 70(1): 1-18.

Lucas, P. and H. Wilting. 2018. *Using planetary boundaries to support national implementation of environment-related SDGs*. The Hague: PBL Netherlands Environmental Assessment Agency.

Lucas, P. L., H. C. Wilting, A. F. Hof, and D. P. van Vuuren. 2020. Allocating planetary boundaries to large economies: Distributional consequences of alternative perspectives on distributive fairness. *Global Environmental Change* 60: 102017.

Lutter, S. and S. Giljum. 2015. *Developing targets for global resource use: Targets for freshwater use*. WP2.4. Vienna. IntRESS project.

Mace, G. M., M. Barrett, N. D. Burgess, S. E. Cornell, R. Freeman, M. Grooten, and A. Purvis. 2018. Aiming higher to bend the curve of biodiversity loss. *Nature Sustainability* 1(9): 448-451.

Mace, G. M., B. Reyers, R. Alkemade, R. Biggs, F. S. Chapin, S. E. Cornell, S. Díaz, S. Jennings, P. Leadley, P. J. Mumby, A. Purvis, R. J. Scholes, A. W. R. Seddon, M. Solan, W. Steffen, and G. Woodward. 2014. Approaches to defining a planetary boundary for biodiversity. *Global Environmental Change* 28: 289-297.

Maes, J., A. Driver, B. Czúcz, H. Keith, B. Jackson, E. Nicholson, and M. Dasoo. 2020. A review of ecosystem condition accounts: lessons learned and options for further development. *One Ecosystem* 5: e53485.

Mair, S., A. Jones, J. Ward, I. Christie, A. Druckman, and F. Lyon. 2018. A Critical Review of the Role of Indicators in Implementing the Sustainable Development Goals. In *Handbook of Sustainability Science and Research*, edited by W. Leal Filho: Springer.

Mannering, J. V. 1981. The use of soil loss tolerances as a strategy for soil conservation. In *Soil Conservation: Problems and Prospects*, edited by R. P. C. Morgan. Cranfield: John Wiley & Sons.

Martín-Fernández, J. A., C. Barceló-Vidal, and V. Pawlowsky-Glahn. 2003. Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. *Mathematical Geology* 35(3): 253-278.

Maxim, L., J. H. Spangenberg, and M. O'Connor. 2009. An analysis of risks for biodiversity under the DPSIR framework. *Ecological Economics* 69(1): 12-23.

Maxwell, S. L., R. A. Fuller, T. M. Brooks, and J. E. Watson. 2016. Biodiversity: The ravages of guns, nets and bulldozers. *Nature* 536(7615): 143-145.

McCormack, D. E., K. K. Young, and L. W. Kimberlin. 1982. Current Criteria for Determining Soil Loss Tolerance. In *Determinants of Soil Loss Tolerance*, edited by American Society of Agronomy and Soil Science Society of Ametica. Madison.

McGlade, C. and P. Ekins. 2015. The geographical distribution of fossil fuels unused when limiting global warming to 2 degrees C. *Nature* 517(7533): 187-U143.

McGowan, P. J. K., G. B. Stewart, G. Long, and M. J. Grainger. 2019. An imperfect vision of indivisibility in the Sustainable Development Goals. *Nature Sustainability* 2(1): 43-45.

McLeod, A. R. and R. A. Skeffington. 1995. The Liphook Forest Fumigation Project: an overview. *Plant, Cell & Environment* 18(3): 327-335.

Mehtälä, J. and T. Vuorisalo. 2007. Conservation policy and the EU Habitats Directive: favourable conservation status as a measure of conservation success. *European Environment* 17(6): 363-375.

Meili, M., K. Bishop, L. Bringmark, K. Johansson, J. Munthe, H. Sverdrup, and W. de Vries. 2003. Critical levels of atmospheric pollution: criteria and concepts for operational modelling of mercury in forest and lake ecosystems. *Science of The Total Environment* 304(1): 83-106.

Meltzer, E. 2009. *The Quest for Sustainable International Fisheries: Regional Efforts to Implement the 1995 United Nations Fish Stocks Agreement : an Overview for the May 2006 Review Conference*. Ottawa: NRC Research Press,.

Millennium Ecosystem Assessment. 2005. *Ecosystems and Human Well-being: Synthesis*. Washington, DC: Island Press.

Mills, G., H. Harmens, F. Hayes, H. Pleijel, P. Büker, and I. González-Fernández. 2017. III. Mapping critical levels for vegetation. In *Manual on methodologies and criteria for Modelling and Mapping Critical Loads & Levels and Air Pollution Effects, Risks and Trends*, edited by UBA. Berlin: Umweltbundesamt.

Mills, G., A. Buse, B. Gimeno, V. Bermejo, M. Holland, L. Emberson, and H. Pleijel. 2007. A synthesis of AOT40-based response functions and critical levels of ozone for agricultural and horticultural crops. *Atmospheric Environment* 41(12): 2630-2643.

Miola, A. and F. Schiltz. 2019. Measuring sustainable development goals performance: How to monitor policy action in the 2030 Agenda implementation? *Ecological Economics* 164: 106373.

Moldan, B., S. Janoušková, and T. Hák. 2012. How to understand and measure environmental sustainability: Indicators and targets. *Ecological Indicators* 17: 4-13.

Moldenhauer, W. C. and C. A. Onstad. 1975. Achieving specified soil loss levels. *Journal of Soil and Water Conservation* 30: 166-168.

Molle, F., E. López-Gunn, and F. v. Steenbergen. 2018. The Local and National Politics of Groundwater Overexploitation. *Water Alternatives* 11(3): 445-457.

Montgomery, D. R. 2007. Soil erosion and agricultural sustainability. *Proceedings of the National Academy of Sciences* 104(33): 13268-13272.

Montzka, S. A., G. S. Dutton, P. Yu, E. Ray, R. W. Portmann, J. S. Daniel, L. Kuijpers, B. D. Hall, D. Mondeel, C. Siso, J. D. Nance, M. Rigby, A. J. Manning, L. Hu, F. Moore, B. R. Miller, and J. W. Elkins. 2018. An unexpected and persistent increase in global emissions of ozone-depleting CFC-11. *Nature* 557(7705): 413-417.

Montzka, S. A., G. S. Dutton, R. W. Portmann, M. P. Chipperfield, S. Davis, W. Feng, A. J. Manning, E. Ray, M. Rigby, B. D. Hall, C. Siso, J. D. Nance, P. B. Krummel, J. Mühle, D. Young, S. O'Doherty, P. K. Salameh, C. M. Harth, R. G. Prinn, R. F. Weiss, J. W. Elkins, H. Walter-Terrinoni, and C. Theodoridi. 2021. A decline in global CFC-11 emissions during 2018−2019. *Nature* 590(7846): 428-432.

Mooney, C., Juliet Eilperin, Desmond Butler, John Muyskens, A. Narayanswamy, and N. Ahmed. 2021. Countries' climate pledges built on flawed data, Post investigation finds. *The Washington Post* [Washington, D.C.], section.

Morgan, R. P. C. 2005. *Soil erosion and conservation*. Third edition ed. Carlton: Blackwell Publishing Ltd.

Mudgal, S., A. Tan, S. Lockwood, N. Eisenmenger, M. Fischer-Kowalski, S. Giljum, and M. Brucker. 2012. *Assessment of resource efficiency indicators and targets*. Final report to the European Commission's DG Environment.

Murray, B. B. R., M. J. B. Zeppel, G. C. Hose, and D. Eamus. 2003. Groundwater-dependent ecosystems in Australia: It's more than just water for rivers. *Ecological Management & Restoration* 4(2): 110-113.

Myastkivskyy, V. 2012. *Assessment and Modeling of Available Forest Resources in Europe for Energy Supply*. Ispra: Joint Research Centre - Institute for Environment and Sustainability.

NCC. 2014. *The State of Natural Capital: Restoring our Natural Assets*. Natural Capital Committee.

Neumann, B., K. Ott, and R. Kenchington. 2017. Strong sustainability in coastal areas: a conceptual interpretation of SDG 14. *Sustainability Science* 12(6): 1019-1035.

Neumayer, E. 2003. *Weak versus Strong Sustainability: Exploring the Limits of Two Opposing Paradigms*. Cheltenham: Edward Elgar.

Newman, P. A., J. S. Daniel, D. W. Waugh, and E. R. Nash. 2007. A new formulation of equivalent effective stratospheric chlorine (EESC). *Atmos. Chem. Phys.* 7(17): 4537-4552.

Newman, P. A., L. D. Oman, A. R. Douglass, E. L. Fleming, S. M. Frith, M. M. Hurwitz, S. R. Kawa, C. H. Jackman, N. A. Krotkov, E. R. Nash, J. E. Nielsen, S. Pawson, R. S. Stolarski, and G. J. M. Velders. 2009. What would have happened to the ozone layer if chlorofluorocarbons (CFCs) had not been regulated? *Atmos. Chem. Phys.* 9(6): 2113-2128.

Nilsson, J. and P. Grennfelt. 1988. *Critical Loads for Sulphur and Nitrogen*. Copenhagen: Nordic Council of Ministers.

Nilsson, M., E. Chisholm, D. Griggs, P. Howden-Chapman, D. McCollum, P. Messerli, B. Neumann, A.-S. Stevance, M. Visbeck, and M. Stafford-Smith. 2018. Mapping interactions between the sustainable development goals: lessons learned and ways forward. *Sustainability Science* 13(6): 1489-1503.

Nykvist, B., Å. Persson, F. Moberg, L. Persson, S. Cornell, and J. Rockström. 2013. *National Environmental Performance on Planetary Boundaries*. Stockholm: Swedish Environmental Protection Agency.

O'Brien, T. G., J. E. M. Baillie, L. Krueger, and M. Cuke. 2010. The Wildlife Picture Index: monitoring top trophic levels. *Animal Conservation* 13(4): 335-343.

O'Brien, M. 2015. Timber consumption and sustainable forest use: Assessing the EU's current and expected consumption of global timber in relation to the global capacity for sustainable supplythesis, Faculty of Civil and Environmental Engineering, University of Kassel, Munich.

O'Brien, M., H. Schütz, and S. Bringezu. 2015. The land footprint of the EU bioeconomy: Monitoring tools, gaps and needs. *Land Use Policy* 47: 235-246.

OECD. 1993. *OECD core set of indicators for environmental performance reviews*. No. 83. Paris: Organisation for Economic Co-operation and Development. Environment monographs.

OECD. 2008. *Measuring Material Flows and Resource Productivity: Volume I. The OECD Guide*. Paris: Organisation for Economic Co-operation and Development.

OECD. 2013. *OECD Compendium of Agri-environmental Indicators*. Paris: Organisation for Economic Co-operation and Development.

OECD. 2019. *Measuring Distance to the SDG Targets 2019: An Assessment of Where OECD Countries Stand*. Paris: organisation for Economic Co-operation and Development.

OECD and JRC. 2008. *Handbook on Constructing Composite Indicators: Mehodology and User Guide*. Paris: Organisation for Economic Co-operation and Development.

Osipova, E., Y. Shi, C. Kormos, P. Shadie, Z. C., and T. Badman. 2014. *IUCN World Heritage Outlook 2014: A conservation assessment of all natural World Heritage sites*. Gland: International Union for the Conservation of Nature.

Osipova, E., M. Emslie-Smith, M. Osti, M. Murai, U. Åberg, and P. Shadie. 2020. *IUCN World Heritage Outlook 3: A conservation assessment of all natural World Heritage sites*. Gland: International Union for the Conservation of Nature.

Osipova, E., P. Shadie, C. Zwahlen, M. Osti, Y. Shi, C. Kormos, B. Bertzky, M. Murai, R. Van Merm, and T. Badman. 2017. *IUCN World Heritage Outlook 2: A conservation assessment of all natural World Heritage sites*. Gland: International Union for the Conservation of Nature.

Otieno, M. N., C. Mbaisi, P. Nguru, O. Mungai, F. Inganga, W. Busenei, A. Omambia, E. Ooko, S. Maithya, A. Lamutete, and C. Soto-Navarro. 2021. *Strong Environmental Sustainability in Kenya. A case study in the Environmental Sustainability Gap Framework (ESGAP) and Strong Environmental Sustainability Index (SESI).* Nairobi: Agence Française de Développement.

Ozone Secretariat United Nations Environment Programme. 2009. *Handbook for the Montreal Protocol on Substances that Deplete the Ozone Layer*. Nairobi: Secretariat for The Vienna Convention for the Protection of the Ozone Layer & The Montreal Protocol on Substances that Deplete the Ozone Layer.

Ozone Secretariat United Nations Environment Programme. 2019. *Data centre*. Secretariat for The Vienna Convention for the Protection of the Ozone Layer & The Montreal Protocol on Substances that Deplete the Ozone Layer.

Panagos, P., C. Ballabio, J. Poesen, E. Lugato, S. Scarpa, L. Montanarella, and P. Borrelli. 2020. A Soil Erosion Indicator for Supporting Agricultural, Environmental and Climate Policies in the European Union. 12(9): 1365.

Panagos, P., P. Borrelli, J. Poesen, C. Ballabio, E. Lugato, K. Meusburger, L. Montanarella, and C. Alewell. 2015. The new assessment of soil loss by water erosion in Europe. *Environmental Science & Policy* 54: 438-447.

Papadimitriou, E., A. R. Neves, and W. Becker. 2019. *JRC Statistical Audit of the Sustainable Development Goals Index and Dashboards*. Ispra: Joint Research Centre.

Papadimitriou, E., A. Fragoso Neves, and M. Saisana. 2020. *JRC Statistical Audit of the 2020 Environmental Performance Index*. Luxembourg: Joint Research Centre.

Pasimeni, P. 2013. The Europe 2020 Index. *Social Indicators Research* 110(2): 613-635.

Pastor, A. V., F. Ludwig, H. Biemans, H. Hoff, and P. Kabat. 2014. Accounting for environmental flow requirements in global water assessments. *Hydrology and Earth System Sciences* 18(12): 5041-5059.

Patrick, M., J. S. Tenywa, P. Ebanyat, M. M. Tenywa, D. N. Mubiru, T. A. Basamba, and A. Leip. 2013. Soil Organic Carbon Thresholds and Nitrogen Management in Tropical Agroecosystems: Concepts and Prospects. *Journal of Sustainable Development* 6(12): 31-43.

Peters, G. P. 2016. The 'best available science' to inform 1.5 [deg]C policy choices. *Nature Clim. Change* advance online publication.

Pickering, J. and Å. Persson. 2020. Democratising planetary boundaries: experts, social values and deliberative risk evaluation in Earth system governance. *Journal of Environmental Policy & Planning* 22(1): 59-71.

Piet, G. J., A. J. Albella, E. Aro, H. Farrugio, J. Lleonart, C. Lordan, B. Mesnil, G. Petrakis, C. Pusch, G. Radu, and H. J. Rätz. 2010. *Marine Strategy Framework Directive Task Group 3 Report: Commercially exploited fish and shellfish*. EUR 24316. Joint Research Centre - Institute for Environment and Sustainability. JRC Scientific and Technical Reports.

Pintér, L., P. Hardi, A. Martinuzzi, and J. Hall. 2012. Bellagio STAMP: Principles for sustainability assessment and measurement. *Ecological Indicators* 17: 20-28.

Poelman, H. 2018. *A walk to the park? Assessing access to green areas in Europe's cities*. WP 01/2018. European Commission.

Punt, A. E. and A. D. M. Smith. 2002. The gospel of maximum sustainable yield in fisheries management: birth, crucifixion and reincarnation. In *Conservation of exploited species*, edited by J. D. Reynolds, et al. Cambridge: Cambridge University Press.

Radermacher, W. J. 2019. Governing-by-the-numbers/Statistical governance: Reflections on the future of official statistics in a digital and globalised society. *Statistical Journal of the IAOS* 35: 519-537.

Randalls, S. 2010. History of the 2°C climate target. *Wiley Interdisciplinary Reviews: Climate Change* 1(4): 598-605.

Raskin, P. D., E. Hansen, and R. M. Margolis. 1996. Water and sustainability. *Nat Resour Forum* 20(1): 1-15.

Raskin, P. D., P. Gleick, P. Kirshen, G. Pontius, and K. Strzepek. 1997. *Comprehensive assessment of the water resources of the world*. Stockholm: Stockholm Environment Institute.

Rees, W. E. and M. Wackernagel. 2013. The Shoe Fits, but the Footprint is Larger than Earth. *PLOS Biology* 11(11): e1001701.

Reeves, D. W. 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. *Soil and Tillage Research* 43(1–2): 131-167.

Richardson, D. H. S. 1988. Understanding the pollution sensitivity of lichens. *Botanical Journal of the Linnean Society* 96(1): 31-43.

Richter, B. D. 2010. Re-thinking environmental flows: from allocations and reserves to sustainability boundaries. *River Research and Applications* 26: 1052–1063.

Richter, B. D., M. M. Davis, C. Apse, and C. Konrad. 2012. A PRESUMPTIVE STANDARD FOR ENVIRONMENTAL FLOW PROTECTION. *River Research and Applications* 28(8): 1312-1321.

Rickels, W., J. Dovern, J. Hoffmann, M. F. Quaas, J. O. Schmidt, and M. Visbeck. 2016. Indicators for monitoring sustainable development goals: An application to oceanic development in the European Union. 4(5): 252-267.

Rigby, M., S. Park, T. Saito, L. M. Western, A. L. Redington, X. Fang, S. Henne, A. J. Manning, R. G. Prinn, G. S. Dutton, P. J. Fraser, A. L. Ganesan, B. D. Hall, C. M. Harth, J. Kim, K. R. Kim, P. B. Krummel, T. Lee, S. Li, Q. Liang, M. F. Lunt, S. A. Montzka, J. Mühle, S. O'Doherty, M. K. Park, S. Reimann, P. K. Salameh, P. Simmonds, R. L. Tunnicliffe, R. F. Weiss, Y. Yokouchi, and D. Young. 2019. Increase in CFC-11 emissions from eastern China based on atmospheric observations. *Nature* 569(7757): 546-550.

Rijsberman, F. R. 2006. Water scarcity: Fact or fiction? *Agricultural Water Management* 80(1-3): 5-22.

Ripple, W. J., C. Wolf, T. M. Newsome, P. Barnard, and W. R. Moomaw. 2020. World Scientists' Warning of a Climate Emergency. *BioScience* 70(1): 8-12.

Ripple, W. J., C. Wolf, T. M. Newsome, M. Galetti, M. Alamgir, E. Crist, M. I. Mahmoud, W. F. Laurance, and s. s. f. countries. 2017. World Scientists' Warning to Humanity: A Second Notice. *BioScience* 67(12): 1026-1028.

Rockström, J., W. Steffen, K. Noone, Å. Persson, I. F. S. Chapin, E. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. Schellnhuber, B. Nykvist, C. A. D. Wit, T. Hughes, S. v. d. Leeuw, H. Rodhe, S. Sörlin, P. K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R. W. Corell, V. J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J. Foley. 2009a. A safe operating space for humanity. *Nature* 461(7263): 472-475.

Rockström, J., W. Steffen, K. Noone, Å. Persson, I. F. S. Chapin, E. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. Schellnhuber, B. Nykvist, C. A. D. Wit, T. Hughes, S. v. d. Leeuw, H. Rodhe, S. Sörlin, P. K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R. W. Corell, V. J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J. Foley. 2009b. Planetary boundaries:exploring the safe operating space for humanity. *Ecology and Society* 14(2).

Rogelj, J., G. Luderer, R. C. Pietzcker, E. Kriegler, M. Schaeffer, V. Krey, and K. Riahi. 2015. Energy system transformations for limiting end-of-century warming to below 1.5 °C. *Nature Climate Change* 5(6): 519-527.

Rogelj, J., M. Schaeffer, P. Friedlingstein, N. P. Gillett, D. P. van Vuuren, K. Riahi, M. Allen, and R. Knutti. 2016. Differences between carbon budget estimates unravelled. *Nature Clim. Change* 6(3): 245-252.

Röschel, L., R. Noebel, U. Stein, S. Naumann, C. Romão, E. Tryfon, Z. Gaudillat, S. Roscher, D. Moser, T. Ellmauer, M. Löhnertz, L. Halada, A. Staneva, and C. Rutherford. 2020. *State of Nature in the EU - Methodological paper Methodologies under the Nature Directives reporting 2013-2018 and analysis for the State of Nature 2000*. Paris: European Topic Centre on Biological Diversity.

Rounsevell, M. D. A., M. Harfoot, P. A. Harrison, T. Newbold, R. D. Gregory, and G. M. Mace. 2020. A biodiversity target based on species extinctions. *Science* 368(6496): 1193.

Running, S. W. 2012. A Measurable Planetary Boundary for the Biosphere. *Science* 337(6101): 1458-1459.

Sachs, J., G. Schmidt-Traub, C. Kroll, G. Lafortune, and G. Fuller. 2019. *Sustainable Development Report 2019*. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN).

Sachs, J., G. Schmidt-Traub, C. Kroll, G. Lafortune, G. Fuller, and F. Woelm. 2020. *The Sustainable Development Goals and COVID-19. Sustainable Development Report 2020*. Cambridge: Cambridge University Press.

Saisana, M. and A. Saltelli. 2011. Rankings and Ratings: Instructions for Use. *Hague Journal on the Rule of Law* 3(2): 247-268.

Saisana, M., A. Saltelli, and S. Tarantola. 2005. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168(2): 307-323.

Samper, C. 2009. Rethinking biodiversity. *Nature Reports Climate Change*(0910): 118-119.

Sánchez, E. 2021. Some 250 kilos of dead fish wash up in Murcia's Mar Menor. *El País,* , section.

Schaefer, M. B. 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Inter-American Tropical Tuna Commission Bulletin* 1(2): 23-56.

Scheffer, M. 2009. *Critical Transitions in Nature and Society*, *Princeton Series in Complexity*. Princeton and Oxford: Princton University Press.

Scheidleder, A. 2012. *Groundwater threshold values: In-depth assessment of the differences in groundwater threshold values established by Member States*. Vienna: Umweltbundesamt.

Schellnhuber, H. J., S. Rahmstorf, and R. Winkelmann. 2016. Why the right climate target was agreed in Paris. *Nature Climate Change* 6: 649.

Schjønning, P., G. Heckrath, and B. T. Christensen. 2009. *Threats to soil quality in Denmark. A review of existing knowledge in the context of the EU Soil Thematic Strategy*. Aarhus: Faculty of Agricultural Sciences - Aarhus University.

Schmidt-Bleek, F. 1993. *The Fossil Makers*. Birkhäuser, Basel, Boston, Berlin.

Schmidt-Bleek, F. 2008. Factor 10: The future of stuff. *Sustainability: Science, Practice, & Policy* 4(1): 1-4.

Schröder, A., L. Persson, and A. M. De Roos. 2005. Direct experimental evidence for alternative stable states: a review. *Oikos* 110(1): 3-19.

Schulze, E.-D., C. Körner, B. E. Law, H. Haberl, and S. Luyssaert. 2012. Large-scale bioenergy from additional harvest of forest biomass is neither sustainable nor greenhouse gas neutral. *GCB Bioenergy* 4(6): 611-616.

Schulze, R., J. Guinée, L. van Oers, R. Alvarenga, J. Dewulf, and J. Drielsma. 2020. Abiotic resource use in life cycle impact assessment—Part II – Linking perspectives and modelling concepts. *Resources, Conservation and Recycling* 155: 104595.

Schyns, J. F., A. Y. Hoekstra, M. J. Booij, R. J. Hogeboom, and M. M. Mekonnen. 2019. Limits to the world's green water resources for food, feed, fiber, timber, and bioenergy. *Proceedings of the National Academy of Sciences* 116(11): 4893-4898.

Scientific Committee on Toxicity, E. a. t. E. 2001. *Opinion on the Proposed health standards for a revised bathing Water Directive expressed at the 23rd CSTEE plenary meeting*. Brussels: European Commission.

Sicherl, P. 1973. TIME-DISTANCE AS A DYNAMIC MEASURE OF DISPARITIES IN SOCIAL AND ECONOMIC DEVELOPMENT*. *Kyklos* 26(3): 559-575.

Sicherl, P. 2011. *New Understanding and Insights from Time-Series Data Based on Two Generic Measures: S-Time-Distance and S-Time-Step*. Working Paper No 44. Paris: Organisation for Economic Co-operation and Development. OECD Statistics Working Papers.

Simsek, Y., W. G. Santika, M. Anisuzzaman, T. Urmee, P. A. Bahri, and R. Escobar. 2020. An analysis of additional energy requirement to meet the sustainable development goals. *Journal of Cleaner Production* 272: 122646.

Sironen, S., J. Seppälä, and P. Leskinen. 2015. Towards more non-compensatory sustainable society index. *Environment, Development and Sustainability* 17(3): 587-621.

Smakhtin, V., C. Revenga, and P. Döll. 2004. A Pilot Global Assessment of Environmental Water Requirements and Scarcity. *Water International* 29(3): 307-317.

Smith, B. D. and M. A. Zeder. 2013. The onset of the Anthropocene. *Anthropocene* 4: 8-13.

Smith, G. 2021. Saudis Dismiss Call to End Oil Spending as 'La La Land' Fantasy. *Bloomberg*, section.

Smith, P., S. J. Davis, F. Creutzig, S. Fuss, J. Minx, B. Gabrielle, E. Kato, R. B. Jackson, A. Cowie, E. Kriegler, D. P. van Vuuren, J. Rogelj, P. Ciais, J. Milne, J. G. Canadell, D. McCollum, G. Peters, R. Andrew, V. Krey, G. Shrestha, P. Friedlingstein, T. Gasser, A. Grübler, W. K. Heidug, M. Jonas, C. D. Jones, F. Kraxner, E. Littleton, J. Lowe, J. R. Moreira, N. Nakicenovic, M. Obersteiner, A. Patwardhan, M. Rogner, E. Rubin, A. Sharifi, A. Torvanger, Y. Yamagata, J. Edmonds, and C. Yongsung. 2015. Biophysical and economic limits to negative CO2 emissions. *Nature Climate Change* 6: 42.

Solomon, S., D. J. Ivy, D. Kinnison, M. J. Mills, R. R. Neely, and A. Schmidt. 2016. Emergence of healing in the Antarctic ozone layer. *Science*.

Spaiser, V., S. Ranganathan, R. B. Swain, and D. J. T. Sumpter. 2017. The sustainable development oxymoron: quantifying and modelling the incompatibility of sustainable development goals. *International Journal of Sustainable Development & World Ecology* 24(6): 457-470.

Sparholt, H. and R. M. Cook. 2010. Sustainable exploitation of temperate fish stocks. *Biology Letters* 6(1): 124-127.

Srebotnjak, T., H. Gerdes, A. Best, and S. Cavalieri. 2009. *Evaluation of Indicators for EU Policy Objectives*. D2.1. IN-STREAM project.

Steffen, W., W. Broadgate, L. Deutsch, O. Gaffney, and C. Ludwig. 2015a. The trajectory of the Anthropocene: The Great Acceleration. *The Anthropocene Review* 2(1): 81-98.

Steffen, W., K. Richardson, J. Rockstrom, S. E. Cornell, I. Fetzer, E. M. Bennett, R. Biggs, S. R. Carpenter, W. de Vries, C. A. de Wit, C. Folke, D. Gerten, J. Heinke, G. M. Mace, L. M. Persson, V. Ramanathan, B. Reyers, and S. Sorlin. 2015b. Planetary boundaries: guiding human development on a changing planet. *Science* 347(6223): 1259855.

Steinberger, J. K. and J. T. Roberts. 2010. From constraint to sufficiency: The decoupling of energy and carbon from human needs, 1975–2005. *Ecological Economics* 70(2): 425-433.

Steinmann, Z. J. N., A. M. Schipper, M. Hauck, S. Giljum, G. Wernet, and M. A. J. Huijbregts. 2017. Resource Footprints are Good Proxies of Environmental Damage. *Environmental Science & Technology* 51(11): 6360-6366.

Stricks, V., F. Hinterberger, and J. Moussa. 2015. *Developing targets for global material use*. WP2.3. Vienna. IntRESS project.

Strosser, P., T. Dworak, P. A. G. Delvaux, M. Berglund, G. Schmidt, J. Mysiak, M. Kossida, I. Iacovides, and V. Ashton. 2012. *Gap Analysis of the Water Scarcity and Droughts Policy in the EU*. Colmar. Final report for the European Commission.

Sullivan, M. J. P.S. L. LewisK. Affum-BaffoeC. CastilhoF. CostaA. C. SanchezC. E. N. EwangoW. HubauB. MarimonA. Monteagudo-MendozaL. QieB. SonkéR. V. MartinezT. R. BakerR. J. W. BrienenT. R. FeldpauschD. GalbraithM. GloorY. MalhiS.-I. AibaM. N. AlexiadesE. C. AlmeidaE. A. de OliveiraE. Á. DávilaP. A. LoayzaA. AndradeS. A. VieiraL. E. O. C. AragãoA. Araujo-MurakamiE. J. M. M. AretsL. ArroyoP. AshtonG. Aymard C.F. B. BaccaroL. F. BaninC. BaralotoP. B. CamargoJ. BarlowJ. BarrosoJ.-F. BastinS. A. BattermanH. BeeckmanS. K. BegneA. C. BennettE. BerenguerN. BerryL. BlancP. BoeckxJ. BogaertD. BonalF. BongersM. BradfordF. Q. BrearleyT. BrncicF. BrownB. BurbanJ. L. CamargoW. CastroC. CéronS. C. RibeiroV. C. MoscosoJ. ChaveE. ChezeauxC. J. ClarkF. C. de SouzaM. CollinsJ. A. ComiskeyF. C. ValverdeM. C. MedinaL. da CostaM. DančákG. C. DargieS. DaviesN. D. CardozoT. de HaullevilleM. B. de MedeirosJ. del Aguila PasquelG. DerroireA. Di FioreJ.-L. DoucetA. DourdainV. DroissartL. F. DuqueR. EkoungoulouF. EliasT. ErwinA. Esquivel-MuelbertS. FausetJ. FerreiraG. F. LlampazoE. FoliA. FordM. GilpinJ. S. HallK. C. HamerA. C. HamiltonD. J. HarrisT. B. HartR. HédlB. HeraultR. HerreraN. HiguchiA. HladikE. H. CoronadoI. Huamantupa-ChuquimacoW. H. HuascoK. J. JefferyE. Jimenez-RojasM. KalamandeenM. N. K. DjuikouoE. KearsleyR. K. UmetsuL. K. KhoT. KilleenK. KitayamaB. KlitgaardA. KochN. LabrièreW. LauranceS. LauranceM. E. LealA. LevesleyA. J. N. LimaJ. LisingoA. P. LopesG. Lopez-GonzalezT. LovejoyJ. C. LovettR. LoweW. E. MagnussonJ. Malumbres-OlarteÂ. G. ManzattoB. H. MarimonA. R. MarshallT. MarthewsS. M. de Almeida ReisC. MaycockK. MelgaçoC. MendozaF. MetaliV. MihindouW. MillikenE. T. A. MitchardP. S. MorandiH. L. MossmanL. NagyH. NascimentoD. NeillR. NilusP. N. VargasW. PalaciosN. P. CamachoJ. PeacockC. PendryM. C. Peñuela MoraG. C. PickavanceJ. PipolyN. PitmanM. PlayfairL. PoorterJ. R. PoulsenA. D. PoulsenR. PreziosiA. PrietoR. B. PrimackH. Ramírez-AnguloJ. ReitsmaM. Réjou-MéchainZ. R. CorreaT. R. de SousaL. R. BayonaA. RoopsindA. RudasE. RutishauserK. Abu SalimR. P. SalomãoJ. SchiettiD. SheilR. C. SilvaJ. S. EspejoC. S. ValeriaM. SilveiraM. Simo-DroissartM. F. SimonJ. SinghY. C. Soto SharevaC. StahlJ. StroppR. SukriT. SunderlandM. SvátekM. D. SwaineV. SwamyH. TaedoumgJ. TalbotJ. TaplinD. TaylorH. ter SteegeJ. TerborghR. ThomasS. C. ThomasA. Torres-LezamaP. UmunayL. V. GamarraG. van der HeijdenP. van der HoutP. van der MeerM. van NieuwstadtH. VerbeeckR. VernimmenA. VicentiniI. C. G. VieiraE. V. TorreJ. VleminckxV. VosO. WangL. J. T. WhiteS. WillcockJ. T. WoodsV. WortelK. YoungR. ZagtL. ZemaghoP. A. ZuidemaJ. A. Zwerts and O. L. Phillips. 2020. Long-term thermal sensitivity of Earth's tropical forests. 368(6493): 869-874.

Sutton, P. 2004. *A Perspective on environmental sustainability? A paper for the Victorian Commissioner for EnvironmentalSustainability*. Melbourne.

Svancara, L. K., R. Brannon J, M. Scott, C. R. Groves, R. F. Noss, and R. L. Pressey. 2005. Policy-driven versus Evidence-based Conservation: A Review of Political Targets and Biological Needs. *BioScience* 55(11): 989-995.

Szestay, K. 1970. The hydrosphere and the human environment: An introduction to some hydrological aspects of international research programmes om problems of the humo emvironment. Paper presented at Proceedings of the Wellington Symposium, Wellington.

Talberth, J. and M. Weisdorf. 2017. Genuine Progress Indicator 2.0: Pilot Accounts for the US, Maryland, and City of Baltimore 2012–2014. *Ecological Economics* 142: 1-11.

Taylor, L. and D. F. Hochuli. 2017. Defining greenspace: Multiple uses across multiple disciplines. *Landscape and Urban Planning* 158: 25-38.

Tharme, R. E. 2003. A global perspective on environmental flow assessment: emerging trends in the development and application of environmental flow methodologies for rivers. *River Research and Applications* 19(5-6): 397-441.

Tilman, D., C. Balzer, J. Hill, and B. L. Befort. 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences* 108(50): 20260-20264.

Tomter, S. M., A. Kuliešis, and T. Gschwantner. 2016. Annual volume increment of the European forests—description and evaluation of the national methods used. *Annals of Forest Science*: 1-8.

Tornero, V. and G. Hanke. 2017. *Potential chemical contaminants in the marine environment: An overview of main contaminant lists*. Luxembourg: Joint Research Centre.

Tornero, V., G. Hanke, A. Ausili, A. Haber, A. Mauffret, A. M. Christensen, A. Oros, A. Serrano, B. McHugh, C. Maggi, D. Bijstra, E. McGovern, E. E. M. Nicolaus, I. Hatzianestis, J. Bellas, J. A. Campillo, J. Lušić, J. Mannio, K. Antoniadis, K. d. Cauwer, K. Kamenova, K. Parmentier, L. Viñas, M. Korsjukov, M. Laht, N. Wessel, S. Dimitrova, T. Porsbring, T. Zalewska, U. Kamman, U. Pirntke, V. Coatu, and V. M. León. 2019. *Marine chemical contaminants – support to the harmonization of MSFD D8 methodological standards: Matrices and threshold values/reference levels for relevant substances*. Luxembourg: Joint Research Centre.

Trung Thang, N., C. Soto-Navarro, H. H. Hanh, D. T. P. Anh, V. T. T. Nga, N. T. Thong, and T. Q. Trung. 2021. *Strong Environmental Sustainability in Vietnam. Piloting the Environmental Sustainability Gap Framework (ESGAP) and Strong Environmental Sustainability Index (SESI)*. Agence Française de Développement.

Tschakert, P. 2015. 1.5°C or 2°C: a conduit's view from the science-policy interface at COP20 in Lima, Peru. *Climate Change Responses* 2(1).

Tsyro, S., W. Aas, S. Solberg, A. Benedictow, H. Fagerli, and T. Scheuschner. 2020. Status of transboundary air pollution in 2018. In *Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components*, edited by H. Fagerli, et al. Oslo: Norwegian Meteorological Institute,.

Turner, R. K. 1993. Sustainability: Principles and Practice. In *Sustainable Environmental Economics and Management: Principles and Practice*, edited by R. K. Turner. New York/London,: Belhaven Press.

UBA. 2004. *Manual on methodologies and criteria for Modelling and Mapping Critical Loads & Levels and Air Pollution Effects, Risks and Trends*. Berlin: Umweltbundesamt.

UCL. 2021a. Recipients of 2020/21 UCL-SJTU Strategic Partner Funds. https://www.ucl.ac.uk/global/news/2021/feb/recipients-202021-ucl-sjtu-strategic-partner-funds. Accessed.

UCL. 2021b. Recipients of third annual UCL-Tohoku Strategic Partner Funds. https://www.ucl.ac.uk/global/news/2021/feb/recipients-third-annual-ucl-tohoku-strategic-partner-funds. Accessed.

Umbach, G. 2020. Of numbers, narratives and challenges: Data as evidence in 21st century policy-making. *Statistical Journal of the IAOS* 36: 1043-1055.

UN. 1987. *Montreal Protocol on Substances that Deplete the Ozone Layer*. Vol. 1522, 1-26369. Montreal: United Nations.

UN. 1992. *United Nations Framework Convention on Climate Change*. United Nations.

UN. 2015a. *Transforming our world: the 2030 Agenda for Sustainable Development* United Nations General Assembly. Resolution adopted by the General Assembly.

UN. 2015b. *The Millennium Development Goals Report 2015*. New York: United Nations.

UN. 2019a. *The Sustainable Development Goals Report 2019*. New York: United Nations.

UN. 2019b. *World Population Prospects 2019*. United Nations.

UN. 2020. *Global indicator framework for the Sustainable Development Goalsand targets of the 2030 Agenda for Sustainable Development* A/RES/71/313, E/CN.3/2018/2, E/CN.3/2019/2, E/CN.3/2020/2. United Nations.

UN Environment. 2019a. *Global Environment Outlook – GEO-6: Healthy Planet, Healthy People*. Nairobi.

UN Environment. 2019b. *Global Environment Outlook – GEO-6: Summary for Policymakers*. Nairobi.

UNDESA. 2021. *System of Environmental-Economic Accounting—Ecosystem Accounting. Final Draft*. New York: United Nations Department of Economic and Social Affairs.

UNDP. 2020. *Human Development Report 2020 - Technical notes*. United Nations Development Programme.

UNECE. 2006. *Report on the workshop on critical levels of ozone: further applying and developing the flux-based concept*. ECE/EB.AIR/WG.1/2006/11. United Nations Economic Commission for Europe.

UNECE. 2015. *Updated Handbook for the 1979 Convention on Long-range Transboundary Air Pollution and its Protocols*. New York and Geneva: United Nations Economic Commission for Europe.

UNEP. 2010. *Environmental effects of ozone depletion and its interactions with climate change: 2010 assessment*. Nairobi: United Nations Environment Programme.

UNEP. 2019. *Measuring Progress: Towards Achieving the Environmental Dimension of the SDGs*. Nairobi: United Nations Environment Programme.

UNEP. 2020. *Emissions Gap Report 2020*. Nairobi: United Nations Environment Programme.

UNEP. 2021. *Measuring Progress: Environment and the SDGs*. Nairobi: United Nations Environment Programme.

UNESCO. 1972. *Convention concerning the Protection of the World Cultural and Natural Heritage*. Paris: S. a. C. O. United Nations Educational.

UNFCCC. 2015. *Adoption of the Paris Agreement*. FCCC/CP/2015/L.9/Rev.1. Paris.

UNFCCC. 2021. *Nationally determined contributions under the Paris Agreement. Synthesis report by the secretariat* Glasgow: United Nations Framework Convention on Climate Change.

UNSD. 2015. *Discussion paper on Principles of Using Quantification to Operationalize the SDGs and Criteria for Indicator Selection*. New York: United Nations Statistics Division.

US EPA. 2003. Chapter 4C: Erosion and Sediment Control. In *National Management Measures to Control Nonpoint Source Pollution from Agriculture*, edited by US EPA. Washington, D.C.: United States Environmental Protection Agency,.

Usubiaga-Liaño, A., G. M. Mace, and P. Ekins. 2019. Limits to agricultural land for retaining acceptable levels of local biodiversity. *Nature Sustainability* 2(6): 491-498.

van den Berg, N. J., H. L. van Soest, A. F. Hof, M. G. J. den Elzen, D. P. van Vuuren, W. Chen, L. Drouet, J. Emmerling, S. Fujimori, N. Höhne, A. C. Kõberle, D. McCollum, R. Schaeffer, S. Shekhar, S. S. Vishwanathan, Z. Vrontisi, and K. Blok. 2019. Implications of various effort-sharing approaches for national carbon budgets and emission pathways. *Climatic Change*.

van den Bergh, J. C. J. M. and F. Grazi. 2014. Ecological Footprint Policy? Land Use as an Environmental Indicator. *Journal of Industrial Ecology* 18(1): 10-19.

van den Bergh, J. C. J. M. and F. Grazi. 2015. Reply to the first systematic response by the Global Footprint Network to criticism: A real debate finally? *Ecological Indicators* 58: 458-463.

van der Plas, F. 2019. Biodiversity and ecosystem functioning in naturally assembled communities. *Biological Reviews* 0(0).

Van Loon, W., G. Hanke, D. Fleet, S. Werner, J. Barry, J. Strand, J. Eriksson, F. Galgani, D. Gräwe, M. Schulz, T. Vlachogianni, M. Press, E. Blidberg, and D. Walvoort. 2020. *A European Threshold*

Value and Assessment Method for Macro Litter on Coastlines. Luxembourg: Joint Research Centre.

van Vuuren, D. and A. Faber. 2009. *Growing within Limits. A Report to the Global Assembly 2009 of the Club of Rome*. Bilthoven: PBL Netherlands Environmental Assessment Agency.

Vea, E. B., M. Ryberg, K. Richardson, and M. Z. Hauschild. 2020. Framework to define environmental sustainability boundaries and a review of current approaches. *Environmental Research Letters* 15(10): 103003.

Verheijen, F. G. A., R. J. A. Jones, R. J. Rickson, and C. J. Smith. 2009. Tolerable versus actual soil erosion rates in Europe. *Earth-Science Reviews* 94(1–4): 23-38.

Victor, D. G. and C. F. Kennel. 2014. Climate policy: Ditch the 2 degrees C warming goal. *Nature* 514(7520): 30-31.

Villarrubia-Gómez, P., S. E. Cornell, and J. Fabres. 2018. Marine plastic pollution as a planetary boundary threat – The drifting piece in the sustainability puzzle. *Marine Policy* 96: 213-220.

von Weizsäcker, E. U., A. B. Lovins, and L. H. Lovins. 1998. *Factor Four: Doubling Wealth - Halving Resource Use : the New Report to the Club of Rome*: Earthscan Publications Limited.

Walenta, J. 2020. Climate risk assessments and science-based targets: A review of emerging private sector climate action tools. *WIREs Climate Change* 11(2): e628.

WBGU. 2009. *Solving the climate dilemma: The budget approach*. Berlin: WBGU.

WHO. 2000. *Air Quality Guidelines for Europe*. No. 91. Geneva: World Health Organisation. WHO Regional Publications, European Series.

WHO. 2003. *Guidelines for safe recreational water environments. Volume 1: Coastal and fresh waters*. Geneva: World Health Organisation.

WHO. 2005. *WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide. Global update 2005. Summary of risk assessment*. Geneva: World Health Organisation.

WHO. 2011. *Guidelines for Drinking-water Quality*. Geneva: World Health Organisation.

WHO. 2018a. *Burden of disease from household air pollution for 2016. Description of method*. Geneva: World Health Organisation.

WHO. 2018b. *WHO recommendations on scientific, analytical and epidemiological developments relevant to the parameters for bathing water quality in the Bathing Water Directive (2006/7/EC)*. Geneva: World Health Organisation.

WHO. 2020. *Percentage of the population using clean and polluting fuels and technologies for cooking*. Geneva: World Health Organisation.

WHO Regional Office for Europe. 2013. *Review of evidence on health aspects of air pollution – REVIHAAP Project*. Bonn: World Health Organization.

WHO Regional Office for Europe. 2017. *Drinking Water Parameter Cooperation Project. Support to the revision of Annex I Council Directive 98/83/EC on the Quality of Water Intended for Human Consumption (Drinking Water Directive)*. Bonn: World Health Organization.

Wiedmann, T., M. Lenzen, L. T. Keyßer, and J. K. Steinberger. 2020. Scientists' warning on affluence. *Nature Communications* 11(1): 3107.

Williams, C. B. 1937. THE USE OF LOGARITHMS IN THE INTERPRETATION OF CERTAIN ENTOMOLOGICAL PROBLEMS. *Annals of Applied Biology* 24(2): 404-414.

Willis. 2021. Dig coal to save the climate': the folly of Cumbria's plans for a new coalmine. *The Guardian*, section.

Woldeamanuel, M., A. Kent, M. McGee, and S. Carvajal. 2020. Walk access to neighbourhood parks: evaluating availability, quality and connectivity. 173(3): 96-107.

Wolf, C., T. Levi, W. J. Ripple, D. A. Zárrate-Charry, and M. G. Betts. 2021. A forest loss report card for the world's protected areas. *Nature Ecology & Evolution* 5(4): 520-529.

Woodward, A. 2019. Brazil's new president has started taking steps towards damaging the 'lungs of the planet'. *Business Insider*, section.

Worm, B., R. Hilborn, J. K. Baum, T. A. Branch, J. S. Collie, C. Costello, M. J. Fogarty, E. A. Fulton, J. A. Hutchings, S. Jennings, O. P. Jensen, H. K. Lotze, P. M. Mace, T. R. McClanahan, C. Minto, S. R. Palumbi, A. M. Parma, D. Ricard, A. A. Rosenberg, R. Watson, and D. Zeller. 2009. Rebuilding Global Fisheries. *Science* 325(5940): 578-585.

Yale University. 2018. *Environmental Performance Index: 2018 Technical Appendix*. New Haven: Yale Center for Environmental Law & Policy.

Zal, N., G. Bariamis, A. Zachos, E. Baltas, and M. Mimikou. 2017. *Use of Freshwater Resources in Europe – An assessment based on water quantity accounts*. 1/2017. Magdeburg: c. a. m. w. European Topic Centre on inland. ETC/ICM Technical Report.

Zeng, Y., S. Maxwell, R. K. Runting, O. Venter, J. E. M. Watson, and L. R. Carrasco. 2020. Environmental destruction not avoided with the Sustainable Development Goals. *Nature Sustainability*.

Annex 1: Supporting information for chapter 4

1. Indicator fiches

Forest resources

General information	
Indicator no.	01
Function	Source functions
Principle	Renew renewable resources
Topic	Biomass
SES indicator	
Indicator	Forest utilization rate
Range	0-∞
Unit	%
Description	The utilization rate is represented as the ratio between fellings and net annual increment, the latter being equal to gross increment minus natural losses.
Data provider	Forest Europe
Data source	Forest Europe (2020)
Time	2015 (*)
Normalisation bounds	gp_{min} = 100, gp_{max} = 70
Notes	(*) The following countries do not report data for net annual increment and/or fellings in 2015: BG, CY, GR, IT, LU, LV, PL, PT and ES. For net annual increment, when countries reported data in 2010, the same 2010-2015 change has been assumed as the net annual increment in O'Brien (2015) (R=0.98). In the case of GR, an interpolation has been undertaken based on the linear relationship between both datasets. For fellings, when countries reported data in 2010, the same 2010-2015 change has been assumed as the roundwood production reported by FAOSTAT (2020) (R=0.98). For LU, the median fellings-to-roundwood production ratio of the sample has been used.
SESP indicator	
Indicator no.	01
Data source	Forest Europe (2020)
Time	t_0= 2010, t_1= 2015
Target value	x_{tr} = 70
Science-based environmental standard(s)	
Indicator	Fellings / Net Annual Increment
Value / Range	70 - 100
Unit	%
Scale	Country
Description	An utilization rate below the standard improves the forest's potential for wood production, and the conditions it provides for biodiversity, health, recreation and other forest functions EEA (2017). An utilization rate above 100 leads to younger forests, lower biomass pools, depleted soil nutrient stocks and a loss of other ecosystem functions (Schulze et al. 2012).
Reference	EEA (2017)

Fish resources

General information

Indicator no.	02
Function	Source functions
Principle	Renew renewable resources
Topic	Biomass

SES indicator

Indicator	Fish stocks within safe biological limits
Range	0-100
Unit	%
Description	The indicator shows the % of commercial fish and shellfish stocks that fall within European jurisdiction that are in good environmental status as defined in the Marine Strategy Framework Directive.
Data provider	European Environment Agency
Data source	EEA (2019d) (*)
Time	2017
Normalisation bounds	gp_{min} = 0, gp_{max} = 100
Notes	(*) In the EU, fish quotas are decided at supranational level, so the all the fish stocks are considered at the same time and the same score is assigned to countries. At the time reported, fish stocks in national UK waters were reported with those of the 27 Member States.

SESP indicator

Notes	Not included

Science-based environmental standard(s)

Indicator	Stock status based on fishing mortality and spawning stock biomass
Value / Range	Fishing mortality and spawning stock biomass consistent with Maximum Sustainable Yield (stock specific).
Unit	Units Tonnes
Scale	Stock
Description	Good environmental status is currently assessed using criteria related to fishing mortality and reproductive capacity. Because of data availability, this information is not always available for all stocks, so sometimes judgements have to be made based on information for fishing mortality or reproductive capacity. There is a third criterion (population age and size distribution) not assessed due to the absence of reference values. (**) The Maximum Sustainable Yield represents the maximum average biomass that can be harvested in the long-term without impeding the remaining stock in fisheries to reproduce itself. Fishing mortality higher the maximum sustainable yield and spawning stock biomass lower than those consistent with the maximum sustainable yield are considered to jeopardise the sustainable long-term exploitation of the fishery and to increase the risk of compromising the recruitment potential of the stock.
Reference	EC (2010)
Notes	(**) ICES recommends an approach based on precautionary mortality and spawning stock biomass. Nonetheless, the Directive uses mortality and spawning stock biomass consistent with maximum sustainable yield as references.

Groundwater resources

General information

Indicator no.	03
Function	Source functions
Principle	Renew renewable resources
Topic	Freshwater

SES indicator

Indicator	Groundwater bodies in good quantitative status
Range	0-100
Unit	%
Description	The indicator shows the % area or number of groundwater bodies that are in good quantitative status as defined in the Water Framework Directive (WFD).
Data provider	European Environment Agency
Data source	EEA (2018b)
Time	2015 (*)
Normalisation bounds	gp_{min} = 0, gp_{max} = 100
Notes	(*) The data has been generated as part of the first and second River Basin Management Plans (RBMPs) of the WFD. Caution is advised when comparing countries and when comparing the first and second RBMPs, as the results are affected by the methods countries have used to collect data and often cannot be compared directly.

SESP indicator

Indicator no.	02
Data source	EEA (2018b)
Time	t_0= 2009, t_1= 2015
Target value	x_{tr} = 100
Notes	(*) Same as above

Science-based environmental standard(s)

Indicator	Quantitative status
Value / Range	Good
Unit	-
Scale	Groundwater body
Description	For a groundwater body to be of good quantitative status each of the following criteria need to be met: • available groundwater resource is not exceeded by the long term annual average rate of abstraction; • no significant diminution of surface water chemistry and/or ecology resulting from anthropogenic water level alteration or change in flow conditions that would lead to failure of environmental quality objectives for any associated surface water bodies; • no significant damage to groundwater dependent terrestrial ecosystems resulting from an anthropogenic water level alteration; • no saline or other intrusions resulting from anthropogenically induced sustained changes in flow direction.
Reference	EC (2009)

Freshwater resources

General information

Indicator no.	04
Function	Source functions
Principle	Renew renewable resources
Topic	Freshwater

SES indicator

Indicator	Freshwater bodies not under water stress
Range	0 - 100
Unit	%
Description	The indicator represents the % of freshwater bodies that is not subject to excessive water consumption at any season. (*)
Data provider	European Environment Agency
Data source	EEA (2018c)
Time	2015 (**)
Normalisation bounds	gp_{min} = 0, gp_{max} = 100
Notes	(*) The indicator is computed quarterly to reflect seasonality. It covers all types of freshwater, namely rivers, lakes, reservoirs and groundwater. Groundwater is also considered in a separate indicator, so there is some overlap between this indicator and the one on the quantitative status of groundwater. (**) The data for CY has been estimated based on the linear relationship between this indicator and the annual Water Exploitation Index + (R=-0.55).

SESP indicator

Indicator no.	03
Data source	EEA (2018c)
Time	t_0= 2010, t_1= 2015
Target value	x_{tr} = 100
Notes	-

Science-based environmental standard(s)

Indicator	Blue water consumption / Mean quarterly flows
Value / Range	20
Unit	%
Scale	(Sub)river basin
Description	Consumption over mean runoff exceeding 20% is commonly used to distinguish water stressed bodies. At this point, the numerator does not subtract environmental flow requirements and therefore excessive consumption does not only reflect the scarcity of the resource, but also its capacity to support freshwater-dependent ecosystems. For this reason, this indicator can only be considered a proxy of freshwater resource scarcity.
Reference	Raskin et al. (1997)

Soil resources

General information

Indicator no.	05
Function	Source functions
Principle	Use non-renewables prudently
Topic	Soil

SES indicator

Indicator	Area with tolerable soil erosion
Range	0 - 100
Unit	%
Description	The indicator shows the % of terrestrial area that is not subject to excessive water soil erosion.
Data provider	European Commission Joint Research Centre
Data source	Panagos et al. (2020)
Time	2016
Normalisation bounds	gp_{min} = 0, gp_{max} = 100

SESP indicator

Indicator no.	04
Data source	Panagos et al. (2020)
Time	t_0= 2010, t_1= 2016
Target value	x_{tr} = 100

Science-based environmental standard(s)

Indicator	Soil erosion rate
Value / Range	1
Unit	t ha^{-1} yr^{-1}
Scale	Local
Description	Rates higher than the reference value lead to loss of agricultural productivity and decrease in water quality.
Reference	Jones et al. (2004); Huber et al. (2008); Verheijen et al. (2009)

Climate change

General information

Indicator no.	06
Function	Sink functions
Principle	Prevent global warming and the depletion of the ozone layer
Topic	Earth System

SES indicator

Indicator	Per-capita CO_2 emissions (*)
Range	0-∞
Unit	t per capita
Description	This indicator shows the per-capita CO_2 emissions of countries (**)
Data provider	Eurostat
Data source	Eurostat (2019a)
Time	2018
Normalisation bounds	gp_{min} = 2.5, gp_{max} = 0.5
Notes	(*) It does not consider international bunkers, and the emissions from land use, land-use change and forestry (LULUCF). The net cumulative contribution of the CO_2 emissions from LULUCF is considered to be close to zero in most 1.5º-2ºC scenarios during the 2010-2100 period (Lucas et al. (2020) based on Clarke et al. (2014)) and are therefore excluded from the country totals. (**) The carbon emission budgets of the Intergovernmental Panel on Climate Change (IPCC 2018) show the 33%, 50% and 67% chances of meeting the 1.5ºC and 2ºC targets only for CO_2 emissions.

SESP indicator

Indicator no.	05
Data source	Eurostat (2019a)
Time	t_0= 2013, t_1= 2018
Target value	x_{tr} = 0.5

Science-based environmental standard(s)

Indicator	Per-capita CO_2 emissions consistent with global climate targets
Value / Range	0.5 – 2.5
Unit	t per capita
Scale	Country
Description	0.5 t per capita are consistent with meeting the 1.5ºC target with 67% of possibilities. On the other hand, 2.5 t per capital is consistent with meeting the 2ºC target with 33% of possibilities. Emissions have been allocated on an equal-per-capita basis using cumulative population figures.
Reference	IPCC (2018); UN (2019b)

Depletion of the ozone layer

General information

Indicator no.	07
Function	Sink functions
Principle	Prevent global warming and the depletion of the ozone layer
Topic	Earth System

SES indicator

Indicator	Per-capita consumption of ODS
Range	(-)∞ - ∞
Unit	t ODP (ozone depleting potential) per capita
Description	This indicator shows the deviation of per-capita consumption (production + import – export (including destruction)) of ODS.
Data provider	Ozone Secretariat United Nations Environment Programme
Data source	Ozone Secretariat United Nations Environment Programme (2019) (*)
Time	2019
Normalisation bounds	gp_{min} = 0.00032; gp_{max} = 0
Notes	(*) The EU is the signatory body to the Montreal Protocol, not the countries. Thus, EU emissions are reported in aggregated form, which results in the same score being assigned to all the countries. UK reports the data as part of the 28-country block.

SESP indicator

Indicator no.	06
Data source	Ozone Secretariat United Nations Environment Programme (2019) (*)
Time	t_0= 2014, t_1= 2019
Target value	x_{tr} = 0
Notes	(*) The EU is the signatory body to the Montreal Protocol, not the countries. Thus, EU emissions are reported in aggregated form, which results in the same score being assigned to all the countries. UK reports the data as part of the 28-country block.

Science-based environmental standard(s)

Indicator	Per-capita consumption of ODS
Value / Range	0 – 0.00032
Unit	t ODP (ozone depleting potential) per capita
Scale	National
Description	The Montreal Protocol is regarded as a key factor behind the early signs of recovery in the Antarctica (Solomon et al. 2016). In principle, long-term country commitments in the Montreal Protocol and its subsequent amendments can be broadly considered environmental standards, but more action is required to decrease the pressure on the ozone layer (EEA 2019a). This is interpreted as the need to reduce ODS consumption to zero (upper bound). We provisionally use the per capita consumption of ODS in 1989 (0.00032 t), which represents the peak of the destruction of the ozone layer, to reflect unsustainable conditions.
Reference	UN (1987); EEA (2019a)

Critical levels in terrestrial ecosystems: ozone

General information

Indicator no.	08
Function	Sink functions
Principle	Respect critical loads for ecosystems
Topic	Terrestrial ecosystems

SES indicator

Indicator	Cropland and forested area exposed to safe ozone levels
Range	0 - 100
Unit	%
Description	The indicator shows the % of cropland and forested area not exposed to critical levels of ozone
Data provider	European Topic Centre on Air Pollution, Transport, Noise and Industrial Pollution
Data source	Horálek et al. (2020)
Time	2017
Normalisation bounds	gp_{min} = 0, gp_{max} = 100

SESP indicator

Indicator no.	07
Data source	Ozone Secretariat United Nations Environment Programme (2019) (*)
Time	t_0= 2012, t_1= 2017
Target value	x_{tr} = 100

Science-based environmental standard(s)

Indicator	AOT40
Value / Range	3 (6000) for cropland 5 (10000) for forested areas
Unit	ppm h (µg m^{-3} h)
Scale	Local
Description	AOT40 gives an indication of accumulated ozone exposure, expressed in µg m^{-3} h, over a threshold of 40 ppb. It is the sum of the differences between hourly concentrations > 80 µg m^{-3} (40 ppb) and 80 µg m^{-3} accumulated over all hourly values measured between 08:00 and 20:00 (Central European Time) between May and July. The environmental standard for cropland is linked to a 5% decrease in yield in wheat. The environmental standard for forested areas is linked to a 5% decrease in biomass.
Reference	Karlsson et al. (2003); Karlsson et al. (2007); Mills et al. (2007)

Critical loads in terrestrial ecosystems: eutrophication and acidification

General information

Indicator no.	09
Function	Sink functions
Principle	Respect critical loads for ecosystems
Topic	Terrestrial ecosystems

SES indicator

Indicator	Ecosystems not exceeding the critical loads of eutrophication and acidification (*) (**)
Range	0 - 100
Unit	%
Description	This indicator represents the % of area-weighted ecosystems not at risk of transgressing the critical loads of eutrophication (modelled as deposition of N) and acidification (modelled as deposition of N and S).
Data provider	European Monitoring and Evaluation Programme
Data source	Tsyro et al. (2020)
Time	2017
Normalisation bounds	gp_{min} = 0, gp_{max} = 100
Notes	(*) The exceedance of critical loads of eutrophication and acidification is reported in separate maps. In line with the chemical status indicators for freshwater and coastal water bodies, a composite indicator with the one-out-all-out rule has been created by spatially aggregating both maps. (**) The indicator has been allocated to the sink function of terrestrial ecosystems, yet it covers both terrestrial and aquatic ecosystems. The acidification and eutrophication effects of N and S compounds are covered in the chemical status of surface waters. Freshwater areas have been excluded during the spatial aggregation in order to avoid overlaps between this indicator and those covering the chemical status of freshwater systems.

SESP indicator

Indicator no.	08
Data source	Tsyro et al. (2020)
Time	t_0= 2005, t_1= 2017
Target value	x_{tr} = 100

Science-based environmental standard(s)

Indicator	Critical load of eutrophication and acidification
Value / Range	Ecosystem specific
Unit	nitrogen eq ha^{-1} yr^{-1} / acid eq ha^{-1} yr^{-1}
Scale	Ecosystem
Description	Critical loads represent the pollutant deposition levels that lead to significant harmful effects on specified sensitive elements of the environment. In the case of nitrogen compounds, they are set considering that an increase availability of nutrients that can affect the composition of species in low-nutrient ecosystems and lead to an increase the nitrate concentrations in water bodies. For acidifying substances, critical loads consider the impacts on flora and fauna resulting from the release of toxic metals such as Al and the leaching of nutrients from soils.
Reference	CLRTAP (2017)

Chemical status of surface water bodies

General information

Indicator no.	10
Function	Sink functions
Principle	Respect critical loads for ecosystems
Topic	Freshwater ecosystems

SES indicator

Indicator	Surface water bodies in good chemical status
Range	0 - 100
Unit	%
Description	The indicator shows the % area or number of surface water bodies that are in good chemical status as defined in the Water Framework Directive (WFD). Rivers have been chosen as the representative body.
Data provider	European Environment Agency
Data source	EEA (2018b)
Time	2015 (*)
Normalisation bounds	gp_{min} = 0, gp_{max} = 100
Notes	(*) The data has been generated as part of the first and second River Basin Management Plans (RBMPs) of the WFD. Caution is advised when comparing countries and when comparing the first and second RBMPs, as the results are affected by the methods countries have used to collect data and often cannot be compared directly.

SESP indicator

Indicator no.	09
Data source	EEA (2018b)
Time	t_0= 2009, t_1= 2015 (*)
Target value	x_{tr} = 100
Notes	(*) Same as above

Science-based environmental standard(s)

Indicator	Chemical status
Value / Range	Good
Unit	-
Scale	Surface water body
Description	Good chemical status means that the concentration of priority substances does not exceed the relevant environmental quality standards specified in the European legislation, which are intended to protect the most sensitive species from direct toxicity, including predators and humans via secondary poisoning. (**)
Reference	European Parliament and European Council (2008b)
Notes	(**) The Directive on Environmental Quality Standards (European Parliament and European Council 2008b) contains the list of substances and standards that are used to assess the chemical status of surface waters. These standards refer to pollutant concentration in waters. Based on guidelines provided by the EU (EC 2011a), countries can establish their own standards for sediment and/or biota, and use them instead of the water-based standards, which can ultimately lead to differences in the standards adopted across countries.

Chemical status of groundwater bodies

General information

Indicator no.	11
Function	Sink functions
Principle	Respect critical loads for ecosystems
Topic	Freshwater ecosystems

SES indicator

Indicator	Groundwater bodies in good chemical status
Range	0 - 100
Unit	%
Description	The indicator shows the % area or number of groundwater bodies that are in good chemical status as defined in the Water Framework Directive (WFD).
Data provider	European Environment Agency
Data source	EEA (2018b)
Time	2015 (*)
Normalisation bounds	gp_{min} = 0, gp_{max} = 100
Notes	(*) The data has been generated as part of the first and second River Basin Management Plans (RBMPs) of the WFD. Caution is advised when comparing countries and when comparing the first and second RBMPs, as the results are affected by the methods countries have used to collect data and often cannot be compared directly.

SESP indicator

Indicator no.	10
Data source	EEA (2018b)
Time	t_0= 2009, t_1= 2015 (*)
Target value	x_{tr} = 100
Notes	(*) Same as above

Science-based environmental standard(s)

Indicator	Chemical status
Value / Range	Good
Unit	-
Scale	Groundwater body
Description	Good groundwater chemical status is achieved when: • there is no sign of saline intrusion in the groundwater body; • the concentrations of pollutants do not exceed those permitted under the applicable groundwater quality standards or threshold values, including those for drinking water protected areas; • the concentrations of pollutants do not result in failure to achieve the environmental objectives of associated surface waters (as specified in the Water Framework Directive), nor in any significant damage to terrestrial ecosystems that depend directly on the groundwater body. (**)
Reference	EC (2009)
Notes	(**) Countries use different threshold values for chemical substances (Scheidleder 2012) and they monitor a different amount of substances (EEA 2018b), which limits the comparability of the country results.

Chemical status of coastal water bodies

General information

Indicator no.	12
Function	Sink functions
Principle	Respect critical loads for ecosystems
Topic	Marine ecosystems

SES indicator

Indicator	Coastal water bodies in good chemical status
Range	0 - 100
Unit	%
Description	The indicator shows the % area or number of coastal water bodies that are in good chemical status as defined in the Water Framework Directive (WFD). It is used as a proxy for marine ecosystems.
Data provider	European Environment Agency
Data source	EEA (2018b)
Time	2015 (*)
Normalisation bounds	gp_{min} = 0, gp_{max} = 100
Notes	(*) The data has been generated as part of the first and second River Basin Management Plans (RBMPs) of the WFD. Caution is advised when comparing countries and when comparing the first and second RBMPs, as the results are affected by the methods countries have used to collect data and often cannot be compared directly.

SESP indicator

Indicator no.	11
Data source	EEA (2018b)
Time	t_0= 2009, t_1= 2015 (*)
Target value	x_{tr} = 100
Notes	(*) Same as above

Science-based environmental standard(s)

Indicator	Chemical status
Value / Range	Good
Unit	-
Scale	Coastal water body
Description	Good chemical status means that the concentration of priority substances does not exceed the relevant environmental quality standards specified in the European legislation, which are intended to protect the most sensitive species from direct toxicity, including predators and humans via secondary poisoning. (**)
Reference	European Parliament and European Council (2008b)
Notes	(**) The Directive on Environmental Quality Standards (European Parliament and European Council 2008b) contains the list of substances and standards that are used to assess the chemical status of surface waters. These standards refer to pollutant concentration in waters. Based on guidelines provided by the European Commission (EC 2011a), countries can establish their own standards for sediment and/or biota, and use them instead of the water-based standards, which can ultimately lead to differences in the standards adopted across countries.

Ecosystem health of terrestrial ecosystems

General information

Indicator no.	13
Function	Life support functions
Principle	Maintain biodiversity and ecosystem health
Topic	Terrestrial ecosystems

SES indicator

Indicator	Terrestrial habitats in favourable conservation status
Range	0 – 100
Unit	%
Description	The indicator shows the number of terrestrial habitats that are in good conservation status as defined in the Habitats Directive. (*)
Data provider	European Environment Agency
Data source	EEA (2020b)
Time	2018
Normalisation bounds	gp_{min} = 0, gp_{max} = 100
Notes	(*) It considers dune habitats; heath and scrub; Sclerophyllous scrub; grasslands; bogs, mires and fens; rocky habitats; and forests. It excludes coastal and freshwater habitats.

SESP indicator

Indicator no.	12
Data source	EEA (2020b)
Time	t_0= 2012, t_1= 2018
Target value	x_{tr} = 100

Science-based environmental standard(s)

Indicator	Conservation status
Value / Range	Favourable
Unit	-
Scale	Ecosystem
Description	The conservation status of a habitat reflects the sum of the influences action on the habitat that may affect its long-term distribution, abundance and quality. The conservation status of a habitat is defined based on range, area, structure and function. Favourable conservation status is achieved when the following conditions are met: • its natural range and areas it covers within that range are stable or increasing; and • the specific structure and functions which are necessary for its long-term maintenance exist and are likely to continue to exist for the foreseeable future; and • the conservation status of its typical species is good. For conservation status to be favourable, the proportion of a habitats reported as 'good' needs to greater than or equal to 75%.
Reference	Röschel et al. (2020)

Ecosystem health of freshwater ecosystems

General information

Indicator no.	14
Function	Life support functions
Principle	Maintain biodiversity and ecosystem health
Topic	Freshwater ecosystems

SES indicator

Indicator	Surface water bodies in good ecological status
Range	0 - 100
Unit	%
Description	The indicator shows the % size or number of surface water bodies that are in good (or high) ecological status as defined in the Water Framework Directive (WFD). Rivers have been chosen as the representative body.
Data provider	European Environment Agency
Data source	EEA (2018b)
Time	2015 (*) (**)
Normalisation bounds	gp_{min} = 0, gp_{max} = 100
Notes	(*) The data has been generated as part of the first and second River Basin Management Plans (RBMPs) of the WFD. Caution is advised when comparing countries and when comparing the first and second RBMPs, as the results are affected by the methods countries have used to collect data and often cannot be compared directly. (**) The data for MT has been estimated based on the linear relationship between this indicator and "coastal water bodies in good ecological status" (R=0.46).

SESP indicator

Indicator no.	13
Data source	EEA (2018b)
Time	t_0= 2009, t_1= 2015 (*)
Target value	x_{tr} = 100
Notes	(*) Same as above

Science-based environmental standard(s)

Indicator	Ecological status
Value / Range	Good
Unit	-
Scale	Surface water body
Description	The ecological status of surface waters (including artificial and heavily modified water bodies) is determined based on biological, physicochemical and hydromorphological criteria. There are no absolute environmental standards applicable across water bodies, so the ecological status is defined based on the extent to which current values deviate from those attributable to undisturbed conditions. (**)
Reference	EC (2003)
Notes	(**) Except for certain chemical substances, there are not hard fixed standards to determine the overall status of water bodies. The WFD provides a normative definition of high and good ecological status. Ultimately, the characterisation of water bodies depends on how countries characterise the undisturbed conditions and on the intercalibration process aimed at ensuring that the high-good and the good-moderate boundaries in all assessment methods

for biological quality elements correspond to comparable levels of ecosystem alteration (EC 2005).

Ecosystem health of coastal ecosystems

General information

Indicator no.	15
Function	Life support functions
Principle	Maintain biodiversity and ecosystem health
Topic	Marine ecosystems

SES indicator

Indicator	Coastal water bodies in good ecological status
Range	0 - 100
Unit	%
Description	The indicator shows the % size or number of coastal water bodies that are in good (or high) ecological status as defined in the Water Framework Directive (WFD).
Data provider	European Environment Agency
Data source	EEA (2018b)
Time	2015 (*)
Normalisation bounds	gp_{min} = 0, gp_{max} = 100
Notes	(*) The data has been generated as part of the first and second River Basin Management Plans (RBMPs) of the WFD. Caution is advised when comparing countries and when comparing the first and second RBMPs, as the results are affected by the methods countries have used to collect data and often cannot be compared directly.

SESP indicator

Indicator no.	14
Data source	EEA (2018b)
Time	t_0= 2009, t_1= 2015 (*)
Target value	x_{tr} = 100
Notes	(*) Same as above

Science-based environmental standard(s)

Indicator	Ecological status
Value / Range	Good
Unit	-
Scale	Surface water body
Description	The ecological status of surface waters (including artificial and heavily modified water bodies) is determined based on biological, physicochemical and hydromorphological criteria. There are no absolute environmental standards applicable across water bodies, so the ecological status is defined based on the extent to which current values deviate from those attributable to undisturbed conditions. (**)
Reference	EC (2003)
Notes	(**) Except for certain chemical substances, there are not hard fixed standards to determine the overall status of water bodies. The WFD provides a normative definition of high and good ecological status. Ultimately, the characterisation of water bodies depends on how countries characterise the undisturbed conditions and on the intercalibration process aimed at ensuring that the high-good and the good-moderate boundaries in all assessment methods for biological quality elements correspond to comparable levels of ecosystem alteration (EC 2005).

Outdoor air pollution

General information

Indicator no.	16
Function	Human health and welfare functions
Principle	Respect standards for human health
Topic	Human health

SES indicator

Indicator	Population exposed to safe levels of particulate matter lower than 2.5 micrometres or less in diameter
Range	0 - 100
Unit	%
Description	The indicator shows the % of population exposed to lower $PM_{2.5}$ levels than the WHO guideline values.
Data provider	European Topic Centre on Air Pollution, Transport, Noise and Industrial Pollution
Data source	Horálek et al. (2020)
Time	2017
Normalisation bounds	$gp_{min} = 0$, $gp_{max} = 100$

SESP indicator

Indicator no.	15
Data source	Horálek et al. (2020)
Time	t_0= 2012, t_1= 2017 (*)
Target value	$x_{tr} = 100$
Notes	-

Science-based environmental standard(s)

Indicator	Average annual $PM_{2.5}$
Value / Range	10
Unit	$\mu g\ m^{-3}$
Scale	Local
Description	The standard refers to the lowest level at which total, cardiopulmonary and lung cancer mortality have been shown to increase with more than 95% confidence in response to long-term exposure to $PM_{2.5}$.
Reference	WHO (2005)

Indoor air pollution

General information

Indicator no.	17
Function	Human health and welfare functions
Principle	Respect standards for human health
Topic	Human health

SES indicator

Indicator	Population using clean fuels and technologies for cooking
Range	(<5) - (>95)
Unit	%
Description	The indicator shows the percentage of population using clean fuels for cooking.
Data provider	World Health Organisation
Data source	WHO (2020)
Time	2018 (*)
Normalisation bounds	gp_{min} = 0, gp_{max} = 100 (**)
Notes	(*) BG lacks data for 2018, so the 2016 value has been assigned. (**) Countries with a value >95 (the maximum assigned by WHO) get a normalised score of 100.

SESP indicator

Indicator no.	16
Data source	Horálek et al. (2020)
Time	t_0= 2013, t_1= 2018
Target value	x_{tr} = 100

Science-based environmental standard(s)

Indicator	Population using clean fuels and technologies for cooking
Value / Range	100
Unit	%
Scale	National
Description	The indicator shows the percentage of population using clean fuels for cooking. Members of a household using polluting fuels (e.g. coal, wood, charcoal, dung, crop residues and kerosene) for cooking are considered to be exposed to harmful levels indoor air pollution independent from age and gender that are several times higher than the 24-h exposure guidelines values proposed by WHO (WHO 2018a).
Reference	WHO (2018a)

Drinking water pollution

General information

Indicator no.	18
Function	Human health and welfare functions
Principle	Respect standards for human health
Topic	Human health

SES indicator

Indicator	Samples that meet the drinking water criteria
Range	0 - 100
Unit	%
Description	The indicator shows the % of samples that meet the drinking water criteria specified in the European legislation
Data provider	European Commission
Data source	EC (2016)
Time	2013 (*)
Normalisation bounds	$gp_{min} = 0$, $gp_{max} = 100$
Notes	(*) HR lacks data for this indicator. The average value of the other countries has been used.

SESP indicator

Indicator no.	17
Data source	EC (2016)
Time	t_0= 2011, t_1= 2013
Target value	$x_{tr} = 100$

Science-based environmental standard(s)

Indicator	Safe drinking criteria
Value / Range	Multiple
Unit	Multiple
Scale	Sample
Description	Environmental standards in the European legislation are in most cases based on the WHO guideline values available at the time and the input from the Commission's Scientific Advisory Committee. The latest evidence calls for a revision of some of these standards. Standards at country level can be more restrictive and cover additional parameters. Drinking water quality is determined based on 48 parameters grouped in three categories: microbiological parameters, chemical parameters and indicator parameters.
Reference	European Council (1998)

Bathing water pollution

General information

Indicator no.	19
Function	Human health and welfare functions
Principle	Conserve landscape and amenity
Topic	Other welfare

SES indicator

Indicator	Recreational water bodies that meet the ‘excellent’ quality criteria
Range	0 - 100
Unit	%
Description	The indicator shows the % of marine and inland water bodies used for recreational uses that meet the reference values in European legislation.
Data provider	European Environment Agency
Data source	EEA (2020c)
Time	2019
Normalisation bounds	gp_{min} = 0, gp_{max} = 100

SESP indicator

Indicator no.	18
Data source	EEA (2020c)
Time	t_0= 2014, t_1= 2019
Target value	x_{tr} = 100

Science-based environmental standard(s)

Indicator	Concentration of Intestinal Enterococci and Escherichia Coli in recreational waters
Value / Range	200 (intestinal enterococci, inland waters), 500 (Escherichia Coli, inland waters), 100 (intestinal enterococci, coastal and transitional waters), 250 (Escherichia Coli, coastal and transitional waters)
Unit	cfu / 100 ml
Scale	Water system
Description	Repeated exposure to those concentrations is associated with 3% of gastrointestinal illness risk and 1% of acute febrile respiratory illness risk.
Reference	EC (2002)

Access to green areas

General information

Indicator no.	20
Function	Human health and welfare functions
Principle	Conserve landscape and amenity
Topic	Other welfare

SES indicator

Indicator	Urban population with nearby green areas
Range	0 - 100
Unit	%
Description	This indicator represents the % of the population that has green urban areas and forests within walking distance.
Data provider	European Commission
Data source	Poelman (2018)
Time	2012
Normalisation bounds	$gp_{min} = 0$, $gp_{max} = 100$

SESP indicator

Notes	Not included

Science-based environmental standard(s)

Indicator	Walking distance
Value / Range	10
Unit	Minutes
Scale	-
Description	Green areas can fulfil a variety of functions. These can range from ecological values to recreational functions, aesthetic value, a role in promoting public health, or more generally enhancing inhabitants' quality of life. Although there is no agreed definition of walking distance, the standard represents access to the functions above.
Reference	Poelman (2018)

Natural and mixed world heritage sites

General information

Indicator no.	21
Function	Human health and welfare functions
Principle	Conserve landscape and amenity
Topic	Other welfare

SES indicator

Indicator	Natural and mixed world heritage sites that have a good conservation outlook
Range	0 - 100
Unit	%
Description	The indicator shows the % of natural and mixed world heritage sites that are considered to have a good conservation outlook.
Data provider	International Union for Conservation of Nature
Data source	Osipova et al. (2020)
Time	2020
Normalisation bounds	gp_{min} = 0, gp_{max} = 100

SESP indicator

Indicator no.	19
Data source	Osipova et al. (2017); Osipova et al. (2020)
Time	t_0= 2017, t_1= 2020
Target value	x_{tr} = 100

Science-based environmental standard(s)

Indicator	Conservation outlook
Value / Range	Good
Unit	-
Scale	Individual sites
Description	Good conservation outlook is defined based on three elements: the current state and trend of values, the threats affecting those values, and the effectiveness of protection and management.
Reference	Osipova et al. (2014)

2. Example of normalisation process

Table 37: SESI normalised values for the European block

Indicator	Indicator value (I_{t1})	Goalpost min (gp_{min})	Goalpost max (gp_{max})	Normalised score (NI)
So_Fo	73%	100%	70%	90
So_Fi	21%	0%	100%	21
So_SW	76%	0%	100%	76
So_GW	91%	0%	100%	91
So_SE	63%	0%	100%	63
Si_GHG	6.0 t per cap	2.5 t per cap	0.5 t per cap	0
Si_ODS	1.2E-7 t per cap	3.2E-4 t per cap	0 t per cap	100
Si_O3	34%	0%	100%	34
Si_EA	21%	0%	100%	21
Si_SW	63%	0%	100%	63
Si_GW	76%	0%	100%	76
Si_CW	71%	0%	100%	71
LS_TE	16%	0%	100%	16
LS_SW	36%	0%	100%	36
LS_CW	57%	0%	100%	57
HW_OP	26%	0%	100%	26
HW_IP	94%	0%	100%	94
HW_DW	99%	0%	100%	99
HW_BW	88%	0%	100%	88
HW_GS	92%	0%	100%	92
HW_WH	33%	0%	100%	33

The normalisation process follows Equation 1.

So_Fo: Forest utilization; So_Fi: Fish stocks; Surface water scarcity, So_GW: Groundwater scarcity; So_SE: Soil erosion; Si_GHG: CO_2 emissions; Si_ODS: Consumption of ODS; Si_O3: Ozone pollution in terrestrial ecosystems; Si_EA: Eutrophication and acidification in terrestrial ecosystems; Si_SW: Chemical pollution in surface waters; Si_GW: Chemical pollution in groundwater; Si_CW: Chemical pollution in coastal waters; LS_TE: Ecological health of terrestrial ecosystems; LS_SW: : Ecological health of surface waters; LS_CW: Ecological health in coastal waters; HW_OP: Outdoor air pollution; HW_IP: Indoor air pollution; HW_DW: Drinking water quality; HW_BW: Quality of bathing waters; HW_GS: Proximity to green spaces; HW_WH: Conservation of World Heritage sites.

3. Treatment of zeros and small values

In order to select the lower bound to re-scale the normalised scores, the consistency between the country rankings obtained after computing SESI with integers in the 1-5 range were compared with the country rankings obtained based the Copeland and Borda methods. The Copeland and Borda methods are non-parametric tools initially used in the political field to rank candidates, but now also being used in the context of composite indicators (JRC 2019).

The Copeland method is a scoring system in which countries compete against each other across the indicators of a set. The performance in each indicator is compared one after the other, and wins and losses are assigned to countries based on that comparison (Al-Sharrah 2010). In the case of SESI, we have used normalised scores of the 21 indicators in the comparison, since the goalpost method allows two countries to have the same normalised score if their original value was outside the upper and lower bounds used in the normalisation process. For instance, if the normalised score in the indicator *i* of country *a* is higher than that of country *b*, country *a* gets *a* win and country *b* a loss. Draws are not considered in this step. The result is the outranking matrix shown in Figure 36 that when interpreted row-wise shows the % of wins of each country.

Figure 36: Copeland outranking matrix based on SES indicators

	AT	BE	BG	HR	CY	CZ	DK	EE	FI	FR	DE	GR	HU	IE	IT	LV	LT	LU	MT	NL	PL	PT	RO	SK	SI	ES	SE	GB
AT	-	.15	.27	.38	.38	.23	.40	.50	.67	.53	.38	.31	.33	.62	.36	.50	.58	.50	.25	.50	.43	.50	.36	.47	.45	.40	.50	.53
BE	.85	-	.53	.65	.50	.71	.82	.93	1.00	.82	.93	.69	.53	.88	.71	.79	.79	.62	.36	.80	.73	.81	.62	.80	.53	.82	.87	.82
BG	.73	.47	-	.65	.53	.53	.56	.76	.89	.83	.61	.71	.69	.88	.61	.65	.76	.60	.38	.67	.53	.76	.76	.75	.76	.78	.72	.83
HR	.62	.35	.35	-	.47	.29	.50	.53	.61	.50	.44	.62	.38	.81	.41	.60	.67	.57	.36	.50	.35	.41	.53	.50	.47	.67	.44	.61
CY	.62	.50	.47	.53	-	.54	.53	.62	.62	.53	.50	.69	.58	.67	.43	.57	.60	.60	.25	.60	.56	.56	.60	.62	.62	.47	.56	.60
CZ	.77	.29	.47	.71	.46	-	.57	.71	1.00	.69	.64	.62	.43	.71	.54	.64	.71	.57	.33	.57	.57	.64	.47	.64	.62	.50	.57	.64
DK	.60	.18	.44	.50	.47	.43	-	.81	.94	.56	.31	.56	.50	.87	.38	.60	.50	.38	.33	.56	.35	.59	.53	.47	.50	.56	.65	.56
EE	.50	.07	.24	.47	.38	.29	.19	-	.80	.38	.27	.44	.21	.67	.31	.33	.40	.36	.21	.47	.25	.25	.41	.29	.40	.25	.36	.44
FI	.33	.00	.11	.39	.38	.00	.06	.20	-	.18	.24	.35	.21	.50	.18	.31	.31	.29	.20	.25	.06	.24	.28	.27	.29	.18	.07	.24
FR	.47	.18	.17	.50	.47	.31	.44	.62	.82	-	.47	.62	.29	.80	.38	.53	.62	.46	.33	.44	.29	.47	.47	.60	.44	.40	.59	.62
DE	.62	.07	.39	.56	.50	.36	.69	.73	.76	.53	-	.47	.40	.75	.41	.64	.64	.46	.29	.67	.50	.41	.53	.40	.50	.41	.67	.59
GR	.69	.31	.29	.38	.31	.38	.44	.56	.65	.38	.53	-	.36	.60	.40	.47	.50	.54	.14	.62	.44	.50	.47	.53	.50	.38	.53	.56
HU	.67	.47	.31	.62	.42	.57	.50	.79	.79	.71	.60	.64	-	.85	.50	.85	.79	.54	.38	.62	.67	.60	.60	.64	.64	.64	.79	.64
IE	.38	.12	.12	.19	.33	.29	.13	.33	.50	.20	.25	.40	.15	-	.07	.43	.47	.38	.20	.36	.12	.19	.31	.21	.29	.13	.33	.20
IT	.64	.29	.39	.59	.57	.46	.62	.69	.82	.62	.59	.60	.50	.93	-	.53	.62	.62	.36	.50	.47	.65	.59	.73	.53	.69	.65	.73
LV	.50	.21	.35	.40	.43	.36	.40	.67	.69	.47	.36	.53	.15	.57	.47	-	.40	.36	.38	.42	.33	.50	.36	.36	.33	.47	.47	.47
LT	.42	.21	.24	.33	.40	.29	.50	.60	.69	.38	.36	.50	.21	.53	.38	.60	-	.42	.38	.38	.27	.44	.40	.29	.38	.44	.53	.44
LU	.50	.38	.40	.43	.40	.43	.62	.64	.71	.54	.54	.46	.46	.62	.38	.64	.58	-	.18	.58	.57	.50	.62	.50	.50	.46	.62	.69
MT	.75	.64	.62	.64	.75	.67	.67	.79	.80	.67	.71	.86	.62	.80	.64	.62	.62	.82	-	.64	.60	.60	.69	.69	.85	.67	.64	.67
NL	.50	.20	.33	.50	.40	.43	.44	.53	.75	.56	.33	.38	.38	.64	.50	.58	.62	.42	.36	-	.31	.47	.40	.27	.43	.38	.60	.56
PL	.57	.27	.47	.65	.44	.43	.65	.75	.94	.71	.50	.56	.33	.88	.53	.67	.73	.43	.40	.69	-	.56	.56	.53	.62	.53	.65	.59
PT	.50	.19	.24	.59	.44	.36	.41	.75	.76	.53	.59	.50	.40	.81	.35	.50	.56	.50	.40	.53	.44	-	.53	.67	.50	.47	.41	.65
RO	.64	.38	.24	.47	.40	.53	.47	.59	.72	.53	.47	.53	.40	.69	.41	.64	.60	.38	.31	.60	.44	.47	-	.44	.53	.41	.67	.65
SK	.53	.20	.25	.50	.38	.36	.53	.71	.73	.40	.60	.47	.36	.79	.27	.64	.71	.50	.31	.73	.47	.33	.56	-	.47	.33	.53	.47
SI	.55	.47	.24	.53	.38	.38	.50	.60	.71	.56	.50	.50	.36	.71	.47	.67	.62	.50	.15	.57	.38	.50	.47	.53	-	.50	.50	.50
ES	.60	.18	.22	.33	.53	.50	.44	.75	.82	.60	.59	.62	.36	.87	.31	.53	.56	.54	.33	.62	.47	.53	.59	.67	.50	-	.65	.69
SE	.50	.13	.28	.56	.44	.43	.35	.64	.93	.41	.33	.47	.21	.67	.35	.53	.47	.38	.36	.40	.35	.59	.33	.47	.50	.35	-	.35
GB	.47	.18	.17	.39	.40	.36	.44	.56	.76	.38	.41	.44	.36	.80	.27	.53	.56	.31	.33	.44	.41	.35	.35	.53	.50	.31	.65	-

Entries above and below the diagonal add up to 1.0. Only comparisons in which both countries had data have been carried out. For instance, comparing country performance related to coastal water bodies was not possible because not all the countries have coastal waters.

In a second step, the outranking scores are converted into marks of +1, 0 and -1 based on a pairwise comparison of countries, where the +1 score is assigned when a country has an outranking score higher than 0.5, which means that it has outperformed the other country in more than half of the indicators. A score of zero is assigned when both countries draw, and a score of -1 is assigned when a country has an outranking score lower than 0.5 (Benini 2019). The countries are then ranked based on these scores. The results are shown in Table 38.

The Borda method is a simpler ranking system in which countries are assigned points depending on their rankings in each indicator. As explained by Becker et al. (2019), for each indicator the country with the highest score gets N points, the second highest score N-1, and so on. After assigning points for all the indicators, the countries are ranked based on their total points. In order to correct for the fact that not all the countries have a value for all the indicators (e.g. coastal waters in landlocked countries), instead the arithmetic average country points for the indicators with valid entries is computed, which is ultimately used to rank countries. The country rankings are shown in Table 38.

Table 38: Copeland and index rankings based on SES indicators

Country	Copeland	Borda	Geo (1)	Geo (2)	Geo (3)	Geo (4)	Geo (5)
AT	8	6	18	19	19	19	19
BE	27	28	28	28	28	28	28
BG	26	27	22	21	21	22	22
HR	14	16	9	7	7	6	6
CY	24.5	21	11	10	9	8	7
CZ	22	22	16	18	20	20	20
DK	18.5	18.5	14	14	14	15	14
EE	3	3	5	5	5	5	5
FI	1.5	1	1	1	1	1	1
FR	9	11	4	4	4	4	4
DE	17	18.5	24	24	24	24	25
GR	11	10	13	13	12	11	11
HU	24.5	25	21	22	23	23	23
IE	1.5	2	2	2	2	2	2
IT	23	24	10	12	13	13	15
LV	4	5	19	17	17	18	18
LT	5	4	15	15	15	14	13
LU	18.5	14	25	26	26	26	26
MT	28	26	26	27	27	27	27
NL	10	9	27	25	25	25	24
PL	21	23	23	23	22	21	21
PT	15	15	8	8	8	9	9
RO	13	17	20	20	18	17	17
SK	12	12	12	11	11	10	10
SI	16	13	17	16	16	16	16
ES	20	20	7	9	10	12	12
SE	6	8	6	6	6	7	8
GB	7	7	3	3	3	3	3

The table shows the country rankings based on the Copeland method and the geometric mean, where the number between parenthesis represents the minimum value used to replace zeros and small values.

Two statistical measures have been used to assess the consistency between the Copeland and Borda rankings and the ranking of the index scores calculated using alternative lower values set for the normalised scores of the SES indicators: the Spearman correlation coefficient and the standard deviation of the net rank differences. The Spearman correlation coefficient is the Pearson correlation coefficient of ranks, and therefore shows the linear correlation between the Copeland and Borda scores on the one hand, and SESI scores on the other. In this context, it should be noted that given that different ranking methods lead to different results (Al-Sharrah 2010), a very high correlation should not necessarily be expected. In a second test, we use the difference between the index score rankings, and the Copeland and Borda ranking and calculate the resulting standard deviation to check the similarity between ranking systems. In this case, the lower the standard deviation, the higher the similarity between the rankings. The results are shown in Table 39.

Table 39: Similarity between ranking systems based on SES indicators

Test	Copeland	Borda	Geo (1)	Geo (2)	Geo (3)	Geo (4)	Geo (5)
Spearman (C)	1.00	0.98	0.56	0.60	0.62	0.63	0.63
Spearman (B)	0.98	1.00	0.53	0.57	0.59	0.59	0.60
StDev_diff (C)	0.00	1.84	7.68	7.31	7.16	7.12	7.06
StDev_diff (B)	1.84	0.00	7.94	7.61	7.49	7.43	7.32

The first row of the table shows the Spearman correlation coefficient between the Copeland method and the other ranking systems. The values in parenthesis indicate the different values to replace zeros and small values. The second row represents the same, but taking the Borda ranking as reference.

The third row shows the standard deviation of the difference between the Copeland and the other rankings. The differences are obtained by subtracting each of the columns in Table 38 from the Copeland rankings in that table. The fourth row does the same, but using the Borda ranking as reference.

The results show that the linear correlation between the Copeland and Borda rankings, and the index rankings based on the geometric mean increases with the integers used to replace zero and small values. The Spearman coefficient shows a relatively strong correlation between the different ranking systems. The third and fourth rows show the standard deviation of the difference of each ranking with the Copeland and Borda rankings. In this case, the lowest standard deviation, which is a sign of similarity between rankings, is the lowest when using the value five to treat zeros and small values. As a result, the value five has been adopted for this purpose.

4. Statistical coherence analysis

A cross-correlation analysis has been used to assess the statistical coherence of SESI. The analysis has been undertaken in two steps. In the first one, the linear correlation between the indicators and their corresponding topics, sustainability principles and topics has been assessed through the Pearson correlation coefficient. Statistical significance has been defined through p-values lower than 0.01. The results are shown in Table 40.

Generally speaking, there are high correlations between the indicators and the topics they represent (e.g. between the forest resources indicator and the score for the biomass topic, to which it contributes) for those indicators in the source, sink and life support functions. The indicators on fish stocks, CO_2 emissions and consumption of ODS do not have a correlation coefficient because all the countries have the same normalised score. This should not occur if the country sample were to include non-European countries. Other indicators such as soil erosion, or those in life support functions have a correlation coefficient of 1 because there is only one indicator in the topic they represent. Correlation between the indicators in the human health and welfare functions, and their corresponding topics is generally lower. In these cases, the scores of the two topics are dominated by the indicators on outdoor air pollution and the conservation status of World Heritage sites. The main reason for this is that the remaining indicators have generally high normalised scores. As a result, the indicators with lower normalised scores have the highest explanatory power because the geometric mean penalises low performances. There are no indicators that are negatively correlated to their respective topic.

The correlations between indicators and principles on the one hand, and indicators and functions on the other, is similar. In most cases, there is a positive correlation between the indicators their corresponding (sub)dimension, although the number of times this happens is lower than at the level of topic. Nonetheless, if significance were to be defined at 0.05 instead of at 0.01 level, the number of indicators with a positive and significant correlation to principles and topics would increase slightly. At index level, only a handful of indicators are positively correlated. Most of these are related to the sink function of terrestrial, freshwater and coastal ecosystems. Given that geometric means are used to aggregate the normalised scores across layers, the final values tend to reflect more accurately those indicators where countries score worse. Likewise, the indicators where countries consistently report good performances (e.g. indoor air pollution, quality of bathing waters) tend to show either low or no correlation with the index scores. The limited correlation shown by some indicators and the index scores is not necessarily problematic. From a statistical point of view, it shows that the information contained in those indicators has been captured in previous layers (where positive correlation was reported with the topic or principle scores), but is not completely reflected in the final index scores. This can be expected when using the geometric mean because it over-represents low performances. From a practical perspective, this has two implications for how the results are used. First, low scores should be interpreted as one or more functions of natural capital being threatened because of the penalisation of low scores by the geometric mean. Second, the index scores should not be interpreted at face value. As shown in the statistical analysis, some of the information contained in the indicators is lost when aggregating scores across layers. Given that there is still a significant positive correlation between most indicators and the function scores, the latter should feature prominently when communicating the results.

Table 40: Correlation between indicators and the corresponding (sub)dimensions of SESI

Indicator	Corresponding topic	Corresponding principle	Corresponding function	Index
So_Fo	0.99 (*)	0.46	0.15	0.21
So_Fi	-	-	-	-
So_SW	0.98 (*)	0.87 (*)	0.81 (*)	0.39
So_GW	0.72 (*)	0.72 (*)	0.75 (*)	0.23
So_SE	1.00 (*)	1.00 (*)	0.88 (*)	0.30
Si_GHG	-	-	-	-
Si_ODS	-	-	-	-
Si_O3	0.78 (*)	0.47	0.44	0.50 (*)
Si_EA	0.94 (*)	0.62 (*)	0.55 (*)	0.64 (*)
Si_SW	0.85 (*)	0.44	0.52 (*)	0.41
Si_GW	0.54 (*)	0.30	0.37	0.48 (*)
Si_CW	1.00 (*)	0.75 (*)	0.81 (*)	0.60 (*)
LS_TE	1.00 (*)	0.55 (*)	0.55 (*)	0.17
LS_SW	1.00 (*)	0.70 (*)	0.70 (*)	0.53 (*)
LS_CW	1.00 (*)	0.88 (*)	0.88 (*)	0.41
HW_OP	0.98 (*)	0.98 (*)	0.86 (*)	0.62 (*)
HW_IP	0.30	0.30	0.40	0.04
HW_DW	0.28	0.28	0.11	0.26
HW_BW	0.10	0.10	0.00	-0.13
HW_GS	0.22	0.22	0.35	-0.02
HW_WH	0.98 (*)	0.98 (*)	0.77 (*)	0.17

The values represent the Pearson correlation coefficients between each indicator and the corresponding topic, sustainability principle, function, and the index as a whole. Correlations that are significant (p value < 0.01) are marked with an asterisk.

So_Fo: Forest utilization; So_Fi: Fish stocks; Surface water scarcity, So_GW: Groundwater scarcity; So_SE: Soil erosion; Si_GHG: CO_2 emissions; Si_ODS: Consumption of ODS; Si_O3: Ozone pollution in terrestrial ecosystems; Si_EA: Eutrophication and acidification in terrestrial ecosystems; Si_SW: Chemical pollution in surface waters; Si_GW: Chemical pollution in groundwater; Si_CW: Chemical pollution in coastal waters; LS_TE: Ecological health of terrestrial ecosystems; LS_SW: : Ecological health of surface waters; LS_CW: Ecological health in coastal waters; HW_OP: Outdoor air pollution; HW_IP: Indoor air pollution; HW_DW: Drinking water quality; HW_BW: Quality of bathing waters; HW_GS: Proximity to green spaces; HW_WH: Conservation of World Heritage sites.

In a second step, a correlation analysis between the upper dimensions of SESI has been carried out (Table 41). The higher levels of the structure of the index show positive correlations between most elements. All the sustainability principles are highly correlated with their respective function. In the case of critical loads, the Pearson coefficient is close to one because the score for the principle related to the disruption of Earth System processes such as climate change and the depletion of the ozone layer is the same for all the European countries. In the case of the maintenance of biodiversity and ecosystem health, the Pearson coefficient is one because only one sustainability principle is assigned to life support functions.

At the index level, all the functions show a positive correlation with the index scores. In the case of the source and human health and welfare functions, this is significant with p values lower than 0.05, but not 0.01. This supports the previous conclusions in which it was argued that the function scores should be shown alongside the index scores when communicating the results.

Table 41: Correlation between the dimensions of the SESI

Principle	Source	Sink	Life support	Human health & other welfare	Index
Ren	0.83 (*)				0.42
NRen	0.88 (*)				0.30
ES		-			-
CL		0.98 (*)			0.85 (*)
B&E			1.00 (*)		0.48 (*)
HH				0.86 (*)	0.59 (*)
L&A				0.79 (*)	0.10

Function	Index
Source	0.42
Sink	0.84 (*)
Life support	0.48 (*)
Human health & other welfare	0.45

The table on the left represents the Pearson correlation coefficients between principles, and function and index scores. The table on the right represents the Pearson correlation coefficient between function and index scores. Correlations that are significant (p value < 0.01) are marked with an asterisk.

Ren: renew renewable resources; NRen: use non-renewables prudently; ES: prevent global warming & ozone depletion; CL: respect critical loads for ecosystems; B&E: maintain biodiversity and ecosystem health; HH: respect standards for human health; L&A: conserve landscape and amenity.

Annex 2: Supporting information for chapter 5

Table 42 provides a worked example of the normalisation process of SESP indicators. The first column makes reference to the table in which the equations used for the normalisation of $R_{o\text{-}c}$ value are described. The remaining columns show the various variables used in equations 5-7.

Table 42: SESPI normalised values for the European block

Norm	Indicator	t_0	t_1	I_{t0}	I_{t1}	NI_{t0}	NI_{t1}	x_{tr}	$trend_{obs}$	$trend_{des}$	$R_{o\text{-}c}$	NI
Table 33	So_Fo	2010	2015	68.60	73.14	100.00	89.52	70	0.91	-0.21	-4.33	0.00
Table 32	So_SW	2010	2015	82.96	75.87	82.96	75.87	100	-1.42	1.61	-0.88	5.94
Table 32	So_GW	2009	2015	87.06	90.79	87.06	90.79	100	0.62	0.61	1.01	100.00
Table 32	So_SE	2010	2016	63.32	63.48	63.32	63.48	100	0.03	2.61	0.01	50.51
Table 33	Si_GHG	2013	2018	7.24	6.70	0.00	0.00	0.5	-0.11	-0.52	0.21	60.40
Table 33	Si_ODS	2014	2019	0.00	0.00	100.00	99.96	0	0.00	0.00	1.00 [a]	100.00
Table 32	Si_O3	2012	2017	21.28	33.80	21.28	33.80	100	2.50	5.09	0.49	74.57
Table 32	Si_EA	2005	2017	16.91	20.95	16.91	20.95	100	0.34	6.08	0.06	52.77
Table 32	Si_SW	2009	2015	71.77	62.98	71.77	62.98	100	-1.46	2.47	-0.59	20.34
Table 32	Si_GW	2009	2015	74.82	75.77	74.82	75.77	100	0.16	1.62	0.10	54.89
Table 32	Si_CW	2009	2015	75.81	70.68	75.81	70.68	100	-0.86	1.95	-0.44	28.11
Table 32	LS_TE	2012	2018	18.21	16.06	18.21	16.06	100	-0.36	7.00	-0.05	47.43
Table 32	LS_SW	2009	2015	35.57	36.37	35.57	36.37	100	0.13	4.24	0.03	51.58
Table 32	LS_CW	2009	2015	59.10	57.20	59.10	57.20	100	-0.32	2.85	-0.11	44.44
Table 32	HW_OP	2012	2017	11.32	25.93	11.32	25.93	100	2.92	5.70	0.51	75.64
Table 32	HW_IP	2013	2018	94.35	94.08	94.35	94.08	95	-0.06	0.08	-0.72	13.96
Table 32	HW_DW	2011	2013	99.46	99.48	99.46	99.48	100	0.01	0.03	1.00	100.00
Table 32	HW_BW	2014	2019	86.43	88.03	86.43	88.03	100	0.32	1.09	0.29	64.69
Table 32	HW_WH	2017	2020	32.35	33.33	32.35	33.33	100	0.33	6.67	0.05	52.45

So_Fo: Forest utilization; Surface water scarcity, So_GW: Groundwater scarcity; So_SE: Soil erosion; Si_GHG: CO_2 emissions; Si_ODS: Consumption of ODS; Si_O3: Ozone pollution in terrestrial ecosystems; Si_EA: Eutrophication and acidification in terrestrial ecosystems; Si_SW: Chemical pollution in surface waters; Si_GW: Chemical pollution in groundwater; Si_CW: Chemical pollution in coastal waters; LS_TE: Ecological health of terrestrial ecosystems; LS_SW: : Ecological health of surface waters; LS_CW: Ecological health in coastal waters; HW_OP: Outdoor air pollution; HW_IP: Indoor air pollution; HW_DW: Drinking water quality; HW_BW: Quality of bathing watersHW_WH: Conservation of World Heritage sites.

[a]: When the normalised values in t_0 and t_1 are higher than 99, negligible changes can alter $R_{o\text{-}c}$ values. In these cases, the trends are largely aligned with meeting the environmental standard by 2030.

Annex 3: Supporting information for chapter 6

The following tables describe in more detail the mapping of indicators to topics and the existence of environmental standards in each case.

Table 43: Mapping of natural capital SDG and SES indicators to environmental and resource areas

Function	Principle	Topic	Subtopic	SESI	UN	SDG Index	OECD	Eurostat
Source	Renew renewable resources	Biomass	Food resources		Proportion of local breeds classified as being at risk of extinction		Proportion of local breeds classified as known being not at risk	
			Forest resources	Forest utilization rate	Forest area as a proportion of total land area	Permanent deforestation	Land area covered by trees Intensity of use of forest resources	Share of forest area
			Fish resources	Fish stocks within safe biological limits	Proportion of fish stocks within biologically sustainable levels	Percentage of fish stocks overexploited or collapsed by EEZ		Estimated trends in fish stock biomass Assessed fish stocks exceeding fishing mortality at maximum sustainable yield
		Freshwater	Surface water resources	Freshwater bodies not under water stress	Level of water stress: freshwater withdrawal as a proportion of available freshwater resources	Freshwater withdrawal as % total renewable water resources	Water stress	Water exploitation index
			Groundwater resources	Groundwater bodies in good quantitative status		Imported groundwater depletion		
	Use non-renewables prudently	Soil	Soil resources	Area with tolerable soil erosion	Proportion of land that is degraded over total land area			Estimated soil erosion by water Soil sealing index Settlement area per capita
		Abiotic raw materials	Fossil fuels					
			Metal ores					
			Non-metallic minerals					

Sink	Prevent global warming, ozone depletion	Earth System	Climate change	CO_2 emissions	Total GHG emissions per year	Energy-related CO_2 emissions per capita		GHG emissions
					Material footprint [a]	Imported CO_2 emissions, technology-adjusted		Mean near surface temperature deviation
					Domestic material consumption [a]			
			Stratospheric ozone depletion	ODS consumption				
	Respect critical levels and loads for ecosystems	Terrestrial ecosystems	Terrestrial pollution	Cropland and forest area exposed to safe ozone levels				
				Terrestrial ecosystems not exceeding the critical loads of eutrophication and acidification				
		Freshwater ecosystems	Surface water pollution	Surface water bodies in good chemical status	Proportion of bodies of water with good ambient water quality		Nutrient balance	Biochemical oxygen demand in rivers
								Phosphate in rivers
								Gross nitrogen balance on agricultural land
								Ammonia emissions from agriculture
			Groundwater pollution	Groundwater water bodies in good chemical status	Proportion of bodies of water with good ambient water quality		Nutrient balance	Nitrate in groundwater
								Gross nitrogen balance on agricultural land
								Ammonia emissions from agriculture

		Marine ecosystems	Marine pollution	Coastal water bodies in good chemical status	Index of coastal eutrophication Plastic debris density			
Life support	Maintain biodiversity and ecosystem health	Terrestrial ecosystems	Ecosystem health	Terrestrial habitats in favourable conservation status	Mountain Green Cover Index		Mountain Green Cover Index	
			Biodiversity		Red List Index	Red List Index of species survival Imported biodiversity threats	Red List Index	Common bird index Common farmland bird index Grassland butterfly index
		Freshwater ecosystems	Ecosystem health	Surface water bodies in good ecological status	Change in the extent of water-related ecosystems over time		Average annual change in water surface	
			Biodiversity		Red List Index	Red List Index of species survival Imported biodiversity threats	Red List Index	
		Marine ecosystems	Ecosystem health	Coastal water bodies in good ecological status	Average marine acidity (pH) measured at agreed suite of representative sampling stations	Ocean Health Index Goal–Clean Waters		Mean ocean acidity
			Biodiversity		Red List Index	Red List Index of species survival Imported biodiversity threats	Red List Index	

<table>
<tr><td rowspan="7">Human health and welfare</td><td rowspan="7">Respect standards for human health</td><td rowspan="7">Human health</td><td rowspan="6">Outdoor air pollution</td><td rowspan="6">Population exposed to safe levels of outdoor air pollutants</td><td rowspan="3">Mortality rate attributed to household and ambient air pollution</td><td>Age-standardised death rate attributable to household air pollution and ambient air pollution</td><td rowspan="3">Age-standardized mortality rate attributed to ambient air pollution</td><td rowspan="6">Exposure to air pollution by particulate matter</td></tr>
<tr><td>Annual mean concentration of particulate matter of less than 2.5 microns of diameter</td></tr>
<tr><td>Production-based SO_2 emissions</td></tr>
<tr><td rowspan="3">Annual mean levels of fine particulate matter in cities</td><td>Imported SO_2 emissions</td><td rowspan="3">Mean population exposure to PM2.5 in metropolitan areas</td></tr>
<tr><td>Nitrogen production footprint</td></tr>
<tr><td>Net imported emissions of reactive nitrogen</td></tr>
<tr><td>Indoor air pollution</td><td>Population using clean fuels and technologies for cooking</td><td>Mortality rate attributed to household and ambient air pollution</td><td>Age-standardised death rate attributable to household air pollution and ambient air pollution</td><td>Proportion of population with primary reliance on clean fuels and technology (%)</td><td></td></tr>
</table>

					Proportion of population with primary reliance on clean fuels and technology	Access to clean fuels & technology for cooking		
			Noise pollution					Population living in households considering that they suffer from noise
			Drinking water pollution	Samples that meet the drinking water criteria	Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene	Population using at least basic drinking water services	Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene	
					Proportion of population using safely managed drinking water services		Population with access to improved drinking water sources	
			Chemical pollution					Consumption of toxic chemicals
	Conserve landscape and amenity	Other welfare	Bathing waters	Recreational water bodies in excellent status				Inland water bathing sites with excellent water quality
								Coastal water bathing sites with excellent water quality
			Green areas	Population with nearby green areas				
			Natural sites	Natural and mixed world heritage sites in good conservation outlook				

Green shading indicates the availability of indicators for those topics; grey shading indicates the absence of such indicators for those topics.

[a]: Domestic Material Consumption and material footprint are part of the environmental SDG indicators, but these represent a proxy of environmental pressures in a broad sense, rather than scarcity as required in the source function.

Table 44: Science-based environmental standards in natural capital SDG and SES indicators

Function	Principle	Topic	Subtopic	SESI	UN	SDG Index	OECD	Eurostat
Source	Renew renewable resources	Biomass	Food resources		Proportion of local breeds classified as being at risk of extinction		Proportion of local breeds classified as known being not at risk	
			Forest resources	Forest utilization rate	Forest area as a proportion of total land area	Permanent deforestation	Land area covered by trees Intensity of use of forest resources	Share of forest area
			Fish resources	Fish stocks within safe biological limits	Proportion of fish stocks within biologically sustainable levels	Percentage of fish stocks overexploited or collapsed by EEZ		Estimated trends in fish stock biomass Assessed fish stocks exceeding fishing mortality at maximum sustainable yield
		Freshwater	Surface water resources	Freshwater bodies not under water stress	Level of water stress: freshwater withdrawal as a proportion of available freshwater resources	Freshwater withdrawal as % total renewable water resources	Water stress	Water exploitation index
			Groundwater resources	Groundwater bodies in good quantitative status		Imported groundwater depletion		
	Use non-renewables prudently	Soil	Soil resources	Area with tolerable soil erosion	Proportion of land that is degraded over total land area			Estimated soil erosion by water Soil sealing index Settlement area per capita
		Abiotic raw materials	Fossil fuels					
			Metal ores					
			Non-metallic minerals					

Sink	Prevent global warming, ozone depletion	Earth System	Climate change	CO_2 emissions	Total GHG emissions per year	Energy-related CO_2 emissions per capita		GHG emissions
					Material footprint [a]	Imported CO_2 emissions, technology-adjusted		Mean near surface temperature deviation
					Domestic material consumption [a]			
			Stratospheric ozone depletion	ODS consumption				
	Respect critical levels and loads for ecosystems	Terrestrial ecosystems	Terrestrial pollution	Cropland and forest area exposed to safe ozone levels				
				Terrestrial ecosystems not exceeding the critical loads of eutrophication and acidification				
		Freshwater ecosystems	Surface water pollution	Surface water bodies in good chemical status	Proportion of bodies of water with good ambient water quality		Nutrient balance	Biochemical oxygen demand in rivers
								Phosphate in rivers
								Gross nitrogen balance on agricultural land
								Ammonia emissions from agriculture
			Groundwater pollution	Groundwater water bodies in good chemical status	Proportion of bodies of water with good ambient water quality		Nutrient balance	Nitrate in groundwater
								Gross nitrogen balance on agricultural land
								Ammonia emissions from agriculture

		Marine ecosystems	Marine pollution	Coastal water bodies in good chemical status	Index of coastal eutrophication Plastic debris density			
Life support	Maintain biodiversity and ecosystem health	Terrestrial ecosystems	Ecosystem health	Terrestrial habitats in favourable conservation status	Mountain Green Cover Index		Mountain Green Cover Index	
			Biodiversity		Red List Index	Red List Index of species survival Imported biodiversity threats	Red List Index	Common bird index Common farmland bird index Grassland butterfly index
		Freshwater ecosystems	Ecosystem health	Surface water bodies in good ecological status	Change in the extent of water-related ecosystems over time		Average annual change in water surface	
			Biodiversity		Red List Index	Red List Index of species survival Imported biodiversity threats	Red List Index	
		Marine ecosystems	Ecosystem health	Coastal water bodies in good ecological status	Average marine acidity (pH) measured at agreed suite of representative sampling stations	Ocean Health Index Goal–Clean Waters		Mean ocean acidity
			Biodiversity		Red List Index	Red List Index of species survival Imported biodiversity threats	Red List Index	

Human health and welfare	Respect standards for human health	Human health	Outdoor air pollution	Population exposed to safe levels of outdoor air pollutants	Mortality rate attributed to household and ambient air pollution	Age-standardised death rate attributable to household air pollution and ambient air pollution	Age-standardized mortality rate attributed to ambient air pollution	Exposure to air pollution by particulate matter
						Annual mean concentration of particulate matter of less than 2.5 microns of diameter		
						Production-based SO_2 emissions		
					Annual mean levels of fine particulate matter in cities	Imported SO_2 emissions	Mean population exposure to PM2.5 in metropolitan areas	
						Nitrogen production footprint		
						Net imported emissions of reactive nitrogen		
			Indoor air pollution	Population using clean fuels and technologies for cooking	Mortality rate attributed to household and ambient air pollution	Age-standardised death rate attributable to household air pollution and ambient air pollution	Proportion of population with primary reliance on clean fuels and technology (%)	

					Proportion of population with primary reliance on clean fuels and technology	Access to clean fuels & technology for cooking		
			Noise pollution					Population living in households considering that they suffer from noise
			Drinking water pollution	Samples that meet the drinking water criteria	Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene	Population using at least basic drinking water services	Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene	
					Proportion of population using safely managed drinking water services		Population with access to improved drinking water sources	
			Chemical pollution					Consumption of toxic chemicals
	Conserve landscape and amenity	Other welfare	Bathing waters	Recreational water bodies in excellent status				Inland water bathing sites with excellent water quality
								Coastal water bathing sites with excellent water quality
			Green areas	Population with nearby green areas				
			Natural sites	Natural and mixed world heritage sites in good conservation outlook				

Green shading indicates the availability of indicators with science-based reference values for those topics; yellow shading means that the reference value of the indicator can be partially considered science-based. Red shading shows that an indicator exists for that topic, but that it does not have a science-based reference value. Grey shading indicates the absence of an indicator for those topics.