



# How does government expenditure impact sustainable development? Studying the multidimensional link between budgets and development gaps

Omar A. Guerrero<sup>1,2</sup>  · Gonzalo Castañeda<sup>3</sup>

Received: 28 April 2021 / Accepted: 23 November 2021  
© The Author(s) 2022

## Abstract

We develop a bottom-up causal framework to study the impact of public spending on high-dimensional and interdependent policy spaces in the context of socioeconomic and environmental development. Using data across 140 countries, we estimate the indicator-country-specific development gaps that will remain open in 2030. We find large heterogeneity in development gaps, and non-linear responses to changes in the total amount of government expenditure. Importantly, our method identifies bounds to how much a gap can be reduced by 2030 through sheer increments in public spending. We show that these structural bottlenecks cannot be addressed through expenditure on the existing government programs, but require novel micro-policies intended to affect behaviors, technologies, and organizational practices. One particular set of bottlenecks that stands out relates to the environmental issues contained in the sustainable development goals 14 and 15.

**Keywords** Sustainable development goals (SDGs) · Complexity · Agent-based models · Public spending

## Introduction

In recent years, a vast literature on the sustainable development goals (SDGs) and the possibility of reaching them by 2030 has emerged. Some of these studies analyze specific SDGs and explore projections of indicators for different micro-policy interventions (e.g., González-Pier et al. 2016; Porciello et al. 2020; Boeren 2019; Sobczak et al. 2021; Mensi and Udenigwe 2021), while others focus on identifying synergies and trade-offs between different SDGs (indicators or targets) (e.g., Fuso Nerini et al. 2019; Lusseau and Mancini 2019; McGowan et al. 2019; Pedercini et al. 2019; Asadikia et al. 2021). This latter approach provides a more holistic evaluation of policy measures attempting to improve

the performance of specific SDGs. A third variant of studies explores how the nature of the relationships between SDGs has changed over time and how likely it is that trade-offs can successfully transform into synergies in the coming years (e.g., Machingura and Lally 2017; Fader et al. 2018; Kroll et al. 2019; Amos and Lydgate 2020; Philippidis et al. 2020). Finally, a fourth set of studies makes use of expert advice or indicator trends to decipher the extent to which the SDGs might achieve the goals set for 2030 (e.g., Luken et al. 2020; Moyer and Hedden 2020; Pradhan et al. 2021; Benedek et al. 2021; Ionescu et al. 2020).

Two major points stand out from this succinct overview: (1) that a systemic perspective—emphasizing interactions among SDGs—is critical for policy evaluation; and (2) that a comprehensive understanding (quantitative) of how budgetary allocations impact SDG performance is almost entirely absent. This paper focuses on the latter point and tries to fill this knowledge gap by developing a modeling framework to study policy prioritization in the context of the SDGs. Akenroye et al. (2018) mention the importance of addressing the problem of policy prioritization and of leveraging existing budget resources for meeting these goals. Such funding frameworks are necessary to analyze pressing questions related to the effectiveness of public funding on existing government programs, for example: Do changes in the size and distribution

---

Handled by Isabel B. Franco, The University of Queensland, Australia.

✉ Omar A. Guerrero  
oguerro@turing.ac.uk

<sup>1</sup> Department of Economics, UCL, London, UK

<sup>2</sup> The Alan Turing Institute, London, UK

<sup>3</sup> Centro de Investigación y Docencia Económica (CIDE), Mexico, Mexico

of the budget (on existing programs) help, effectively, to close development gaps?<sup>1</sup> What are the most and least sensitive SDGs to such budgetary rearrangements? Can the commitments to the 2030 Agenda be met when there is enough government spending to guarantee successful policies? To what extent do structural factors hinder the effectiveness of existing programs? From the perspective of governments, understanding how their expenditure actions translate, at a systemic level, into effective policies is critical to guarantee the success of any international development agenda.

In this paper, we develop a bottom-up computational model in which public expenditure generates development advancement (with various degrees of effectiveness). A bottom-up approach to budgetary prioritization is necessary to properly account for political–economy factors that are present in a multidimensional and interdependent policy space (Guerrero and Castañeda 2020a; Castañeda et al. 2018). One of the analytic benefits of this agent-computing model is the ability of calibration using coarse-grained data<sup>2</sup> of individual countries without needing to pool cross-national data.<sup>3</sup> We exploit this feature to study the sensitivity of country-specific indicators to changes in public expenditure.

We study the feasibility of the SDGs across 140 countries using data from the 2020 edition of the Sustainable Development Report (SDR) (Sachs et al. 2020).<sup>4</sup> Our three main results are the following. First, we provide estimates of the SDG gaps that might remain by 2030 if government programs were to be kept unaltered.<sup>5</sup> Second, we demonstrate that the sensitivity of these gaps vary—in diverse and non-linear ways across countries and indicators—according to the amount of per capita government expenditure. Third,

we identify the maximum reduction that can be achieved for the SDG gaps by 2030 through sheer expenditure increments. That is to say, there are stringent ‘budgetary frontiers’ that cannot be overcome without addressing long-term structural factors (redesigning the government programs). Altogether, our results provide quantitative and theoretically sound insights into what makes the SDGs unfeasible from the perspective of government expenditure and existing development strategies.<sup>6</sup>

We structure the remainder of the paper in five more sections and an appendix. In Sect. 2, we present the methods employed: model description, network estimation, calibration procedure, goodness of fit, and definition of SDG gaps. In Sect. 3, we describe the sources of our database, which includes time series for development indicators and government expenditure, and explain how we geographically cluster the information for producing visualizations. Then, in Sect. 4, we show different figures describing the main results from our simulations. Section 5 compares alternative methodologies (data-fitting and aggregated models) with our bottom-up computational approach in the context of systemic policy analysis and budgetary allocations. Finally, in Sect. 6, we finalize the paper with a brief summary of the model’s purpose and assumptions, and with a suggestion on how to use the simulation results for country-specific policy guidelines.

## Methods

Essentially, the proposed model is designed to study how different budgetary allocations affect the—simultaneous and interdependent—evolution of a large set of development indicators. The model takes as inputs a vector with initial conditions for the indicators, a network with their interdependencies, a budget size, the fraction of positive changes in the indicators (as a measure of variation), and the final values they achieved in the last period of the sample. With this information, the parameters are calibrated to (1) match the simulated and empirical indicators in their final observations, and (2) match the fraction of positive changes. Due to the interdependent dynamics produced by the model, calibrating its parameters is not trivial. Nonetheless, we devise an efficient method that yields a goodness of fit above 90% for most countries. Our model is a variant of Castañeda et al. (2018) and Guerrero and Castañeda

<sup>1</sup> Informally, a development gap is the distance between the level of development of a policy issue and the goal to be achieved by a government or society. Section 4.1 provides a refinement of this concept through the idea of SDG gaps, which are the expected distance that an indicator will have in 2030 with respect to its goal.

<sup>2</sup> In the agent-computing literature, this type of indicator data are considered coarse grained, since they do not provide disaggregated information about the individual behaviors of the agents: something typically needed to calibrate these models (e.g., microdata or administrative records).

<sup>3</sup> While some studies use non-pooled country-level data to describe the structure of trade-offs and synergies (see Pradhan et al. 2017 and references), their capacity to produce quantitative prospective analysis is limited because proper statistical power can only be achieved—under traditional statistical tools—with a large number of observations (which can only be obtained by pooling cross-country data).

<sup>4</sup> The SDR is produced by the Sustainable Development Solutions Network and the Bertelsmann Stiftung.

<sup>5</sup> A government program is the set of policies that a government has in place to affect a specific development issue. Funding or defunding these programs is a short-/medium-term decision, while redesigning them is a long-term one (which needs to address structural factors). Our model focuses on the former—short/medium-term decisions—so it is assumed that the specific policies in place remain unchanged.

<sup>6</sup> Although the third type of literature mentioned above argues for the need for structural changes to achieve SDGs and to break away from trade-offs, the meaning of structural bottleneck remains broad and often ambiguous in terms of policy instruments. Our paper sheds new light by introducing a more nuanced concept of bottlenecks, one with a direct link to government programs that can be directly affected through budgetary readjustments.

(2020a), with the improvement of accounting for the size of a government's budget. Similar models have been successfully applied to study ex ante policy evaluation (Castañeda and Guerrero 2019a), policy resilience (Castañeda and Guerrero 2018), policy coherence (Guerrero and Castañeda 2020b), public governance (Guerrero and Castañeda 2021), and sub-national development (Guerrero et al. 2021). Some of them have also been used in the provision of policy advice (Castañeda and Guerrero 2019b, d, c; Sulmont et al. 2021; Gobierno del Estado de México 2020). While the full details of the model are provided in Appendix A, here we explain the mechanisms that are most salient for this study and elaborate on the new calibration procedure.

## Model description

The model consists of an agent representing the government or central authority in charge of deciding how to spend a budget of size  $B$ . There are  $N$  policy issues, each one with a level of development measured by an indicator. From these policy issues,  $n \leq N$  can be directly impacted through existing government programs, and we assume that there is one program for each one of them. We call these types of policy issues *instrumental*, while the remaining  $N - n$  are considered *collateral*. An issue may be collateral because there does not exist a policy instrument to directly intervene it, and this may occur because the issue is too aggregate (e.g., GDP growth).<sup>7</sup>

In addition to the government agent, there are  $n$  policy-making agents (functionaries), one in charge of each instrumental indicator<sup>8</sup>. Thus, the problem of the central authority is to allocate  $B$  resources across  $n$  policymakers to improve the  $N$  indicators. Policymakers, however, may have different goals from those of the central authority or may just be inefficient. Therefore, some of the allocated resources might end up diverted or wasted. Let us denote the allocation to instrumental policy issue  $i$  as  $P_i$ , and the amount of resources that the policymaker uses effectively as  $C_i$ ; we say that the latter is the contribution of the policymaker.

In an iterative process, the government agent reallocates its resources, prioritizing the most laggard<sup>9</sup> and the most efficient policy issues. In parallel, the policymakers

try to maximize their benefit by determining a level of  $C_i$  that shows proficiency to the central authority (for political reputation), but that also benefits them through the wasted resources  $P_i - C_i$ . The determination of  $C_i$  happens through a behavioral model of reinforcement learning (which has extensive empirical validation), subjected to the monitoring of the government and to the corresponding penalties in case it spots inefficiencies.

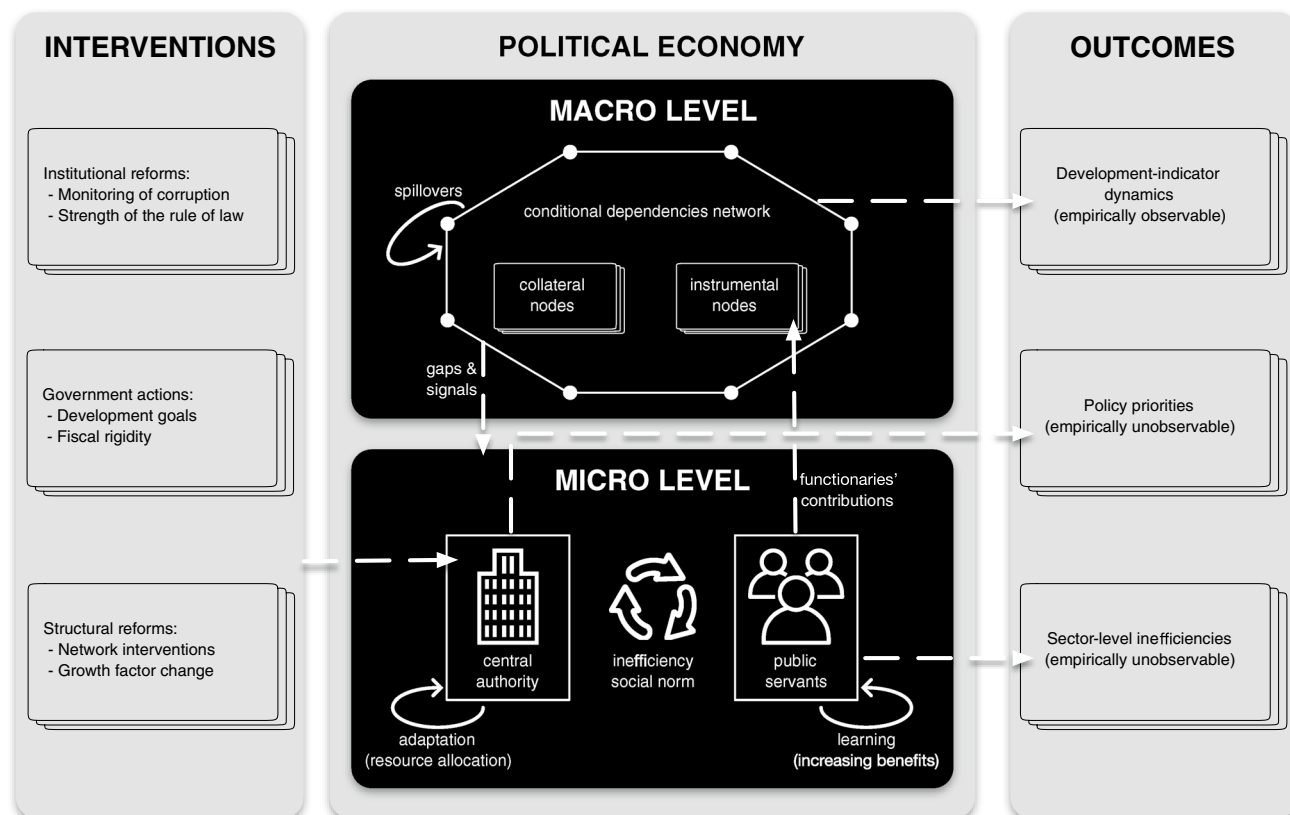
The quality of the procurement mechanisms aimed at minimizing inefficiencies vary across countries according to empirical data on public governance, which we use as an input. With each step, the contributions and the total incoming spillovers  $S_i$  (which could be positive or negative) determine the success probability  $\gamma_i$  of the policy aimed at issue  $i$ . If the policy succeeds, the amount of improvement reflected in the indicator is proportional to the existing long-term structural factors, which we capture explicitly in a parameter  $\alpha_i$ . Altogether, the model runs for  $T$  periods that can be mapped into calendar time. Parameter  $B$  corresponds to the empirical budget that a given country spent during the sampling period. Thus, in the calibration, the budget runs out after  $T$  simulation periods, reflecting different spending capabilities across countries, and enabling the test of potential effects from budgetary increments and reductions. We perform Monte Carlo simulations to generate stable measures of the indicators and other variables of interest. The reader should be aware that the model is calibrated and implemented for each country independently, so this approach overcomes concerns about biases from grouping countries or indicators.

In the interest of clarity and space, we summarize the model in Algorithm 1 and Figure 1. In this section, we focus on the two equations that drive the dynamics of the indicators and provide the details of the remaining equations in Appendix A. These equations connect the outcomes of the behavioral components with the spillover effects shaped by the network of interdependencies, and establish a clear differentiation between short-/mid-term and long-term dynamics.

<sup>7</sup> Other reasons include lack of capacity (e.g., cybersecurity), advanced level of development (e.g., extreme poverty in some advanced economies), or lack of awareness (e.g., pollution and over-exploitation of natural resources in several poor countries).

<sup>8</sup> The model is flexible to accommodate multiple agents per indicator or multiple indicators per agent. This, however, requires detailed contextual information that we leave for country-specific studies.

<sup>9</sup> Prioritizing laggard issues has been a promoted practice since the Millennium Development Project under the assumption that laggard indicators reveal potential bottlenecks.



**Fig. 1** Structure of the model. *Notes:* The left panel shows examples of policy interventions that could be simulated by manipulating some of the model exogenous variables. All the interventions take place at the micro-level and exert a direct impact on budgetary decisions. The panel at the center shows that the model establishes linkages between the micro and the macro levels. At the micro-level, the central authority allocates budgetary resources, while policymakers implement the government programs. At the macro-level, the network of interdependencies produces spillover effects that condition the evolution of the development indicators. In the upward causation component (right-vertical arrow), functionaries make an effective use of some of the resources that they receive from the central government. In the

downward causation (left-vertical arrow), the overall dynamic produces reductions in the development gaps of the 2030 Agenda. This channel also transmits signals reflecting certain misuse of resources, which causes the government to penalize inefficient functionaries and reallocate resources. Moreover, the three circling arrows in the middle of the bottom panel describe a horizontal causation mechanism responsible for the social norms of inefficiency guiding functionaries' behavior. Finally, the left panel presents some of the outcomes that can be obtained from the model: the evolution of the indicators, policy priorities (allocation profiles), and sectoral inefficiencies. *Sources:* Guerrero and Castañeda (2020a)

**Algorithm 1:** Model pseudocode

---

```

1 foreach period  $t$  do
2   foreach public servant  $i$  do
3     receive public funds  $P_{i,t}$ ;
4     evaluate the benefits from the previous contribution  $C_{i,t-1}$ ;
5     establish new contribution level  $C_{i,t}$ ;
6   foreach indicator  $i$  do
7     if the indicator is instrumental then
8       implement public policy using the resources  $C_{i,t}$ ;
9     receive the incoming spillovers  $S_{i,t}$ ;
10    determine the probability of success  $\gamma_{i,t}$  according to  $C_{i,t}$  and  $S_{i,t}$ ;
11    if the public policy is successful (with probability  $\gamma_{i,t}$ ) then
12      improve the indicator according to the long-term structural factors  $\alpha_i$ ;
13  the government monitors the policymakers through imperfect mechanisms;
14  the government penalizes those who are found being inefficient;
15  the policymakers receive the benefit from their chosen contributions;
16  the government updates the allocation profile  $P_{1,t}, \dots, P_{n,t}$ ;

```

---

Now, let us define the evolution of indicator  $i$  as

$$I_{i,t+1} = I_{i,t} + \alpha_i \xi(\gamma_{i,t}), \quad (1)$$

where parameter  $\alpha_i > 0$  captures long-term structural factors.<sup>10</sup> Parameter  $\alpha_i$  imposes a limit to the growth that could be achieved in the short term through sheer spending. For instance, let  $\xi(\gamma_{i,t})$  in equation 1 denote the outcome of a Bernoulli trial that can take values 1 (successful) or 0 (unsuccessful). This means that if a positive event materializes, the indicator grows according to  $\alpha_i$ . As previously mentioned, the probability of a successful trial is  $\gamma_{i,t}$ . Note that  $\gamma_{i,t}$  is an endogenous variable of the model, so we proceed to explain how it is formed.

Recall that the total budget size across periods is  $\mathbf{B}$ . This stock can be turned into flows by defining a disbursement schedule  $B_1, \dots, B_T$ , such that  $\sum_t^T B_t = \mathbf{B}$ . For simplicity, let us assume that the disbursement schedule is homogeneous, so  $B_t = B \quad \forall t$ . Next, consider the allocation profile  $P_{1,t}, \dots, P_{n,t}$  that the central authority defines in period  $t$ . Under the homogeneous disbursement schedule assumption,  $\sum_i^n P_{i,t} = B$  holds, so the contributions of the policymakers are in the same units as the budget. To map  $C_{i,t}$  into the success probability  $\gamma_{i,t}$ , we define

$$\gamma_{i,t} = \beta \frac{C_{i,t} + \frac{1}{n} \sum_j C_{j,t}}{1 + e^{-S_{i,t}}}, \quad (2)$$

<sup>10</sup> Note that if the indicator exceeds its theoretical maximum (if provided by the user), the model will assign zero growth.

where  $\beta$  is a normalizing constant<sup>11</sup> and  $S_{i,t}$  is the total amount of spillovers received by indicator  $i$  in period  $t$  (this could be positive or negative).<sup>12</sup> The spillovers are computed every simulation period according to  $S_{i,t} = \sum_j \mathbf{1}_{j,t} \mathbb{A}_{j,i}$ , where  $\mathbf{1}$  is an indicator function that returns 1 if indicator  $j$  grew in the previous period and 0 otherwise. The adjacency matrix  $\mathbb{A}$  corresponds to the empirical network of interlinkages, with each entry representing a conditional dependence from indicator  $j$  to  $i$ . Importantly, these conditional dependencies do not represent causal links, but rather an empirical regularity that the model takes into account (see Ospina-Forero et al. 2020 for a detailed discussion on estimating SDG networks and the impossibility of interpreting them as causal networks). While the structure of the network represented by  $\mathbb{A}$  is considered a long-term feature, the actual realization of the spillovers is a short-/mid-term phenomenon because it is the result of the dynamics of the other indicators in the previous period.

<sup>11</sup> Importantly, if expenditure data at the level of each indicator were available, it could be used as an input for  $P_i$ , in which case  $\beta_i$  could be indicator-specific and more intuitive in terms of returns to expenditure in specific policy issues. Hence, while we use aggregate expenditure data in this paper, the model is flexible to allow various types of disaggregated data.

<sup>12</sup> The term  $C_{i,t}$  accounts for the expenditure contribution to an instrumental policy issue. For a collateral issue,  $C_{i,t}$  equals zero, so its success depends on the overall ‘financial health’ of the government  $\frac{1}{n} \sum_j C_{j,t}$ , and on the spillovers  $S_{i,t}$ . Therefore, we assume that public funding is a necessary but not sufficient condition for development.

Equation 2 represents the short-/mid-term component of the model, while parameter  $\alpha_i$  from equation 1 captures the long-term factors limiting the impact of public expenditure on the indicators. For example, a government may increase the funds allocated to train quantum-computing engineers with the aim of strengthening this strategic area. While the number of engineers in this field may indeed increase due to the availability of scholarships, they may leave for another country or end up in unrelated jobs due to a lack of employment opportunities in the domestic labor market. A labor market-related structural factor, the demand for quantum-computing engineers, limits the speed with which this sector can develop; such speed will be reflected in modest improvements of the relevant indicators. Naturally, a structural reform could be seen as a change in  $\alpha_i$ , but its interpretation proves difficult due to the multiple variables that are absorbed in this parameter; this is a challenge that we leave for future work. Nevertheless,  $\alpha_i$  is informative about the limits of sheer spending at the level of each indicator, something lacking in all other approaches. For this reason, the model is consistent with the idea of analyzing budgetary changes over existing government programs.

## Networks

As we have previously explained, the structure of the interdependencies between indicators is assumed to be a long-term feature, so the networks are exogenous inputs. As such, adjacency matrices can be built for each country by following any preferred criteria. A popular approach among development scholars is the qualitative approach of eliciting expert opinions. Unfortunately, this strategy is not scalable for a large set of countries and indicators (and is difficult to use in the case governments have severe time constraints). Ospina-Forero et al. (2020) provide a comprehensive review of quantitative methods that may be suitable for estimating SDG networks.<sup>13</sup> With this information in hand, our method of choice is the Bayesian approach of sparse Gaussian Bayesian networks developed by Aragam et al. (2019) (and known as *sparsebn*). This procedure has the distinctive advantages of working well with high-dimensional datasets, even if they have short series, and producing adjacency matrices that try to minimize the number of links that may be false positives (hence the “sparse” term in the name).<sup>14</sup>

Recall that the resulting networks should not be interpreted as causal relations, but as conditional probabilities, which means that a link  $A \rightarrow B$  does not imply that  $\Delta A$  guarantees  $\Delta B$ . This is the reason why spillovers

affect the probability of success  $\gamma_i$ , and not the magnitude of the outcome. Of course, like with any statistical method, *sparsebn* makes certain assumptions such as a linear Gaussian structural equation model and no temporal dependence between observations. The former is a standard assumption in causal Bayesian models. Temporal dependencies can be partially removed by computing the first differences of the series. Overall, we consider that these assumptions are more reasonable than those made by alternative network estimation methods, and further arguments are provided by Ospina-Forero et al. 2020. Finally, the networks are estimated for each country individually, an important improvement over the existing literature on SDG synergies and trade-offs which tends to use pooled data.<sup>15</sup>

## Calibration

The aim of the calibration method is twofold, to assure (1) that the simulated dynamics of the indicators start and end at the empirical levels, and (2) that the model’s average success probability corresponds to the empirical fraction of positive first differences of the indicators.<sup>16</sup> To achieve this, we need to find the parameter vector  $\alpha_1, \dots, \alpha_N, \beta$  that minimizes an error measure.

There are two features that characterize this calibration problem. First, the dynamics of the indicators are interdependent. This means that if  $\alpha_i$  changes, the ‘speed’ of another indicator  $j$  may be altered as well. Furthermore, these interdependencies are not obvious enough so that the model could be written as a system of equations to be simultaneously solved (as one may think by looking at equation 1). For instance, the fact that  $\gamma_i$  is endogenous renders homogeneous Markov chains ineffective. The second feature is the computational cost of each evaluation. Since each simulation may yield a different trajectory for the same indicator, stable metrics have to be obtained from Monte Carlo simulations. This means that evaluating

<sup>13</sup> For instance, correlation thresholding is one of the methods commonly used in the literature (e.g., Warchold et al. 2021; Putra et al. 2020).

<sup>14</sup> For more details on the network and its estimation procedure, see Appendix D.

<sup>15</sup> Naturally, the network plays a role in the model, so different topologies may influence some of the model’s variables. In fact, Castañeda et al. (2018) and Guerrero and Castañeda (2020a) show that removing the spillovers alters the incentive structure of the policymaking agents, resulting in lower variation of inefficiency across policy issues. Nevertheless, for the variables of interest of this study (the SDG gaps), we find that our results are robust to different networks. Appendix I provides detailed evidence.

<sup>16</sup> Appendix E discusses how to deal with indicators that show final values that are lower than their initial conditions.

a given set of parameters involves several independent runs.<sup>17</sup>

We develop a multi-objective gradient descent method that exploits the fact that each parameter can be associated to a specific error. Let us define an indicator-specific error as  $e_{\alpha_i} = I_{i,-1} - \bar{I}_{i,T}$ , where  $I_{i,-1}$  is the final empirical value of indicator  $i$ , and  $\bar{I}_{i,T}$  is the average final simulated value of the same indicator across  $M$  independent Monte Carlo simulations. The corresponding error for  $\beta$  is  $e_{\beta} = \Gamma - \frac{\sum_{i,m} Y_{i,m}}{M \times T \times N}$ , where  $\Gamma$  is the fraction of positive first differences across all indicators. The calibration algorithm tries to minimize the average absolute error  $e = \frac{1}{N+1} \left( \sum_i |e_{\alpha_i}| + |e_{\beta}| \right)$ .

To minimize the error, first, we start with a proposed vector  $\alpha_1, \dots, \alpha_N, \beta$ . Next, we perform a set of  $M$  Monte Carlo simulations and compute the error vector  $e_{\alpha_1}, \dots, e_{\alpha_N}, e_{\beta}$ . For each indicator  $i$ , if  $e_{\alpha_i} < 0$  (meaning that the indicator grew too fast), then we multiply  $\alpha_i$  by a factor  $1 - \delta_{\alpha_i}$ . If

$e_{\alpha_i} > 0$  (the indicator was too slow), then we multiply  $\alpha_i$  by  $1 + \delta_{\alpha_i}$ . The same logic applies to  $\beta$ , which has a corresponding factor  $\delta_{\beta}$ . Ideally, we want that the mean error converges to zero as we search the parameter space. We can generate such behavior by setting factors  $\delta_{\alpha_1}, \dots, \delta_{\alpha_N}, \delta_{\beta}$  that change in proportion to the errors. As it turns out, a factor that achieves this for indicator  $i$  is  $\delta_{\alpha_i} = |e_{\alpha_i}| / (I_{i,-1} - I_{i,0})$ , where  $I_{i,0}$  and  $I_{i,-1}$  are the empirical initial and final values of the indicator, while  $\delta_{\beta} = |e_{\beta}|$  for  $\beta$ . Our simulations suggest zero-error convergence for a large enough  $M$ .<sup>18</sup> Thus, it can be run for several iterations until a certain threshold for the average error is achieved. The calibration procedure for the model parameters is described in Algorithm 2. As the reader will notice, we bound the step factor  $(1 \pm \delta)$  by  $1/2$  or  $3/2$  as we have found that this accelerates the convergence rate significantly.

---

#### Algorithm 2: Calibration pseudocode

---

```

1 initialize vector  $\alpha_1, \dots, \alpha_N, \beta$  with random values;
2 for an error tolerance threshold do
3   run  $M$  Monte Carlo simulations;
4   compute the error vector  $e_{\alpha_1}, \dots, e_{\alpha_N}, e_{\beta}$ ;
5   foreach indicator  $i$  do
6     if  $e_{\alpha_i} < 0$  then
7       update  $\alpha_i$  to  $\alpha_i \times \max(1 - |\delta_{\alpha_i}|, 1/2)$ ;
8     else
9       update  $\alpha_i$  to  $\alpha_i \times \min(1 + |\delta_{\alpha_i}|, 3/2)$ ;
10  if  $e_{\beta} < 0$  then
11    update  $\beta$  to  $\beta \times \max(1 - |\delta_{\beta}|, 1/2)$ ;
12  else
13    update  $\beta$  to  $\beta \times \min(1 + |\delta_{\beta}|, 3/2)$ ;

```

---

<sup>17</sup> Heuristic optimization algorithms that can handle dynamic landscapes, such as simulated annealing and particle swarm fail, arguably due to the sensitivity of the fitness landscape and to the cost of each evaluation. Evolutionary approaches such as differential evolution have also been ineffective due to similar reasons. Finally, Bayesian methods, such as the tree-structured Parzen estimator, which perform well with expensive-evaluation models, do not work in this context due to the high dimensionality of the solution space (and the sensitivity of the fitness landscape).

<sup>18</sup> The resulting parameter vector is robust across different calibrations using random initial parameters.

## Goodness of fit

For a single indicator  $i$ , the goodness of fit of its corresponding parameter  $\alpha_i$  is

$$\text{GoF}_{\alpha_i} = 1 - \frac{e_{\alpha_i}}{I_{i,-1} - I_{i,0}}, \quad (3)$$

which takes values in the interval  $(-\eta, 1]$ , where  $\eta$  is the lower bound induced by the theoretical maximum of the indicator. If no theoretical maximum exists, then the lower bound is  $-\text{inf}$ .

The basic idea behind  $\text{GoF}_{\alpha_i}$  is that, in a good fit, the error  $e_{\alpha_i}$  should represent a small fraction of the historical gap that needs to be closed in a simulation ( $I_{i,-1} - I_{i,0}$ ). Errors where the simulated average indicator ends below the empirical value are bound by  $I_{i,0}$  because the model only allows non-negative growth. However, an error where the simulated average indicator ends above the empirical value may represent multiple times the size of the historical gap. Therefore, this metric not only takes into account accuracy with respect to the final value, but it also penalizes extreme errors with negative contributions when computing the mean goodness of fit across all indicators. Importantly, when testing alternative calibration methods, several indicators display a negative  $\text{GoF}_{\alpha_i}$ . This is not the case for our algorithm.

The metric for the goodness of fit of parameter  $\beta$  follows the same logic, but the target feature is the rate of positive first differences. Formally, the goodness of fit of  $\beta$  is

$$\text{GoF}_{\beta} = 1 - \frac{e_{\beta}}{\Gamma}, \quad (4)$$

where  $\Gamma$  is the number of positive first differences in the empirical data as a fraction of all first differences.

The overall goodness of fit for a country is the average

$$\text{GoF} = \frac{1}{N+1} \left( \sum_i^N \text{GoF}_{\alpha_i} + \text{GoF}_{\beta} \right). \quad (5)$$

Figure 2 shows the distribution of the GoF after calibrating the model.<sup>19</sup> More detailed results on the goodness of fit are provided in Appendix F. Notice that, when performing this calibration procedure, we obtain a remarkable goodness of fit at the country level. Furthermore, the large majority of the parameters  $\alpha_i$  exhibit a fitting above 0.9, while this is always the case for  $\beta$ .

<sup>19</sup> The choice of the number of simulation periods  $T$  does not alter the results significantly because the calibration of  $\beta$  compensates for a higher or lower frequency of the disbursement schedule. Appendix H provides evidence of robustness under different disbursement schedules.

## Definition of SDG gaps

The main estimates of the paper are the gaps or the distances between development goals and the levels predicted for the indicators in 2030. If a prediction surpasses its goal, then we say that the gap has been closed. Formally, an SDG gap is

$$\text{gap}_i = \begin{cases} 100 \times \frac{G_i - \bar{I}_{i,T}}{G_i} & G_i \geq \bar{I}_{i,T}, \\ 0 & G_i < \bar{I}_{i,T} \end{cases}, \quad (6)$$

where  $G_1, \dots, G_N$  are the development goals obtained from the SDR (the values that each indicator is supposed to achieve in 2030), and  $\bar{I}_{i,T}$  is the expected value of indicator  $i$ —across  $M$  independent Monte Carlo simulations—after  $T$  simulation periods that are equivalent to 10 years. The underlying yearly budget for the 2021–2030 period is assumed to be identical, in per capita terms, to the (annual) average expenditure observed in the 21 years of data. We express the gaps as a proportion of their goals and in percentage terms. Thus, we can read an SDG gap as: “by 2030, indicator  $i$  will still need to close  $x\%$  of its goal”.<sup>20</sup>

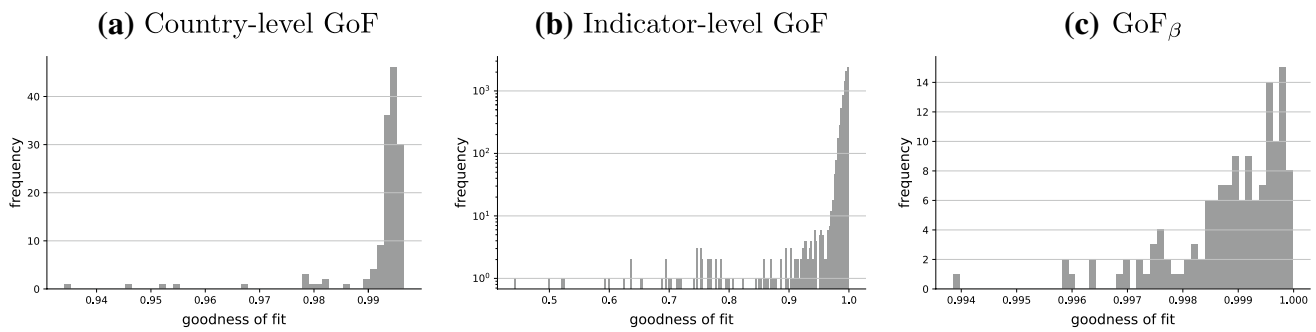
## Data

There exist different databases from which one could obtain indicators classified into the SDGs, for example, the SDG indicators from the United Nations Statistics Division (United Nations 2020), the World Bank Atlas of Sustainable Development Goals (World Bank 2020), the OECD SDG distance indicators (OECD 2020), and the indicators compiled by the Bertelsmann Stiftung and Sustainable Development Solutions Network to produce the Sustainable Development Report (SDR) (Sachs et al. 2020). In this study, we use the SDR database for three main reasons.<sup>21</sup> First, the SDR is the only dataset that provides quantitative values for the goals to be achieved by each indicator. Furthermore, these goals are consistent across all the countries in the sample because the chosen indicators are applicable to each nation. Since the aim of this paper is to assess the

<sup>20</sup> Appendix C reports confidence intervals and provides a method to incorporate uncertainty about the quality of the data into the intervals when information on the indicators’ errors is available.

<sup>21</sup> A caveat of the chosen data is that they do not contain time series for SDG 12 (Responsible consumption and production). The indicators in SDG 12 relate to issues such as waste management, which have just recently been quantified in a handful of countries. However, this is also an issue in all alternative datasets.





**Fig. 2** Distribution of goodness of fit metrics. *Sources:* Authors' own calculations

feasibility of reaching the SDGs, having quantitative goals is necessary. Second, the SDR data have consistently longer time series than alternative databases. This is helpful for the calibration of the model, because the estimation of the structural factors  $\alpha_1, \dots, \alpha_N$  assumes that they capture long-term features of the data. For a sub-sample of 140 countries, the SDR provides time series with a length of almost 21 years (from 2000 to 2020) in numerous indicators. Alternative datasets, while they contain more indicators, fail to provide consistently long time series. Third, the majority of the data sources for the SDR indicators are recognized international (and intergovernmental) organizations, while the rest are scientifically sound products such as surveys from statistics bureaus, NGOs, and academic institutions.

While the SDR team makes a substantial effort in gathering as much data as possible for each country, there are countries that lack some of the indicators, or that have too few observations. For this reason, different countries in our sample may have more or fewer indicators than others. This is problematic for all studies that pool cross-national data, since decisions have to be made regarding the imputation of missing observations, or the complete elimination of certain indicators. Our approach overcomes this problem because we do not need to produce estimates on pooled data. Thus, we allow each country to have its potentially unique set of indicators and perform the estimations independently of other nations.<sup>22</sup> This allows capturing as many policy dimensions as possible for each country, which is consistent with the philosophy behind multidimensional development.

<sup>22</sup> In Appendix D.3, we present a methodology for the imputation of missing observations that works very well when indicators exhibit non-linear dynamics and the network is estimated with pooled country data. This is one of the several methods available for the imputation of missing information in the SDGs. For instance, Gaussian processes are reliable for non-linear dynamics when a database only includes time series for one country (or region), while in cross-sectional analyses heuristic approaches are more common (e.g., Warchoł et al. 2021).

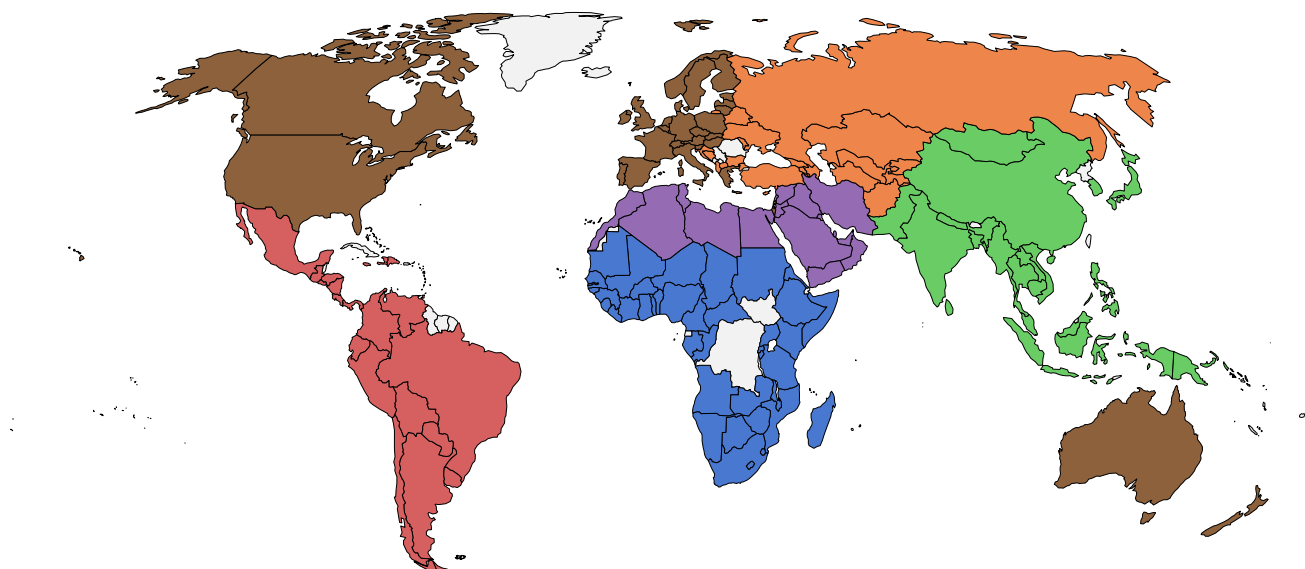
While having unbalanced panels is still not the ideal setup for ex post cross-country comparisons, we believe that this framework is still able to overcome some of the main hurdles of data-fitting approaches. Appendix B provides detailed information on the 77 indicators of our sample, and their distribution across countries.

For the purpose of visualizing some of our results, we may color or aggregate them into country clusters. We should emphasize that this is only for visualization purposes. For these country clusters, we use the following grouping scheme: sub-Saharan Africa (*Africa*), Eastern Europe and Central Asia (*E. Europe & C. Asia*), East and South Asia (*East & South Asia*), Latin America and the Caribbean (*LAC*), Middle East and North Africa (*MENA*), and Western countries (*West*). Figure 3 provides a map of the countries covered in our sample.

For the national budgets, we use data on total government expenditure in current USD (which can be accessed through the link: [data.worldbank.org/indicator/NE.CON.GOVV.KD](https://data.worldbank.org/indicator/NE.CON.GOVV.KD)). This information is obtained from the dataset on General Government Final Consumption Expenditure which, in turn, sources the information from the World Bank National Accounts Data and the OECD National Accounts data files. We compute the total expenditure exercised by each country in the 2000–2020 period. Missing values are imputed with the average yearly expenditure, and the final amount is transformed into per capita expenditure to remove population-size effects (we use the population size reported by the SDR).

## Results

Because we only have aggregate yearly government expenditure for making worldwide comparisons, we limit our analysis to three types of simulation exercises. Firstly, we study whether SDG gaps of the 2030 United Nations Agenda can be closed assuming a benchmark scenario in which we project the historical yearly average of public expenditure



**Fig. 3** Countries and their clusters. *Notes:* Blue: Africa. Orange: E. Europe & C. Asia. Green: East & South Asia. Red: LAC. Purple: MENA. Brown: West. Countries in gray were excluded from the sample due to lack of data

for the following 10 years. Secondly, we analyze the sensitivity of these gaps to different increments/decrements of the budget size. We also visualize the response function of budgetary changes in terms of delays (or savings) in the number of years to reach the 2030 levels obtained in the benchmark scenario. Thirdly, we study structural bottlenecks that hamper the possibility of improving the indicators' performances by increasing the allocated funds. These bottlenecks are made evident when, by construction, inefficiencies and budgetary constraints are ruled out in a counter-factual simulation. Although these exercises are produced at the country level using all available indicators separately, for exposition reasons, we present several visualizations at the SDG or geographical cluster level in the main body of the paper.

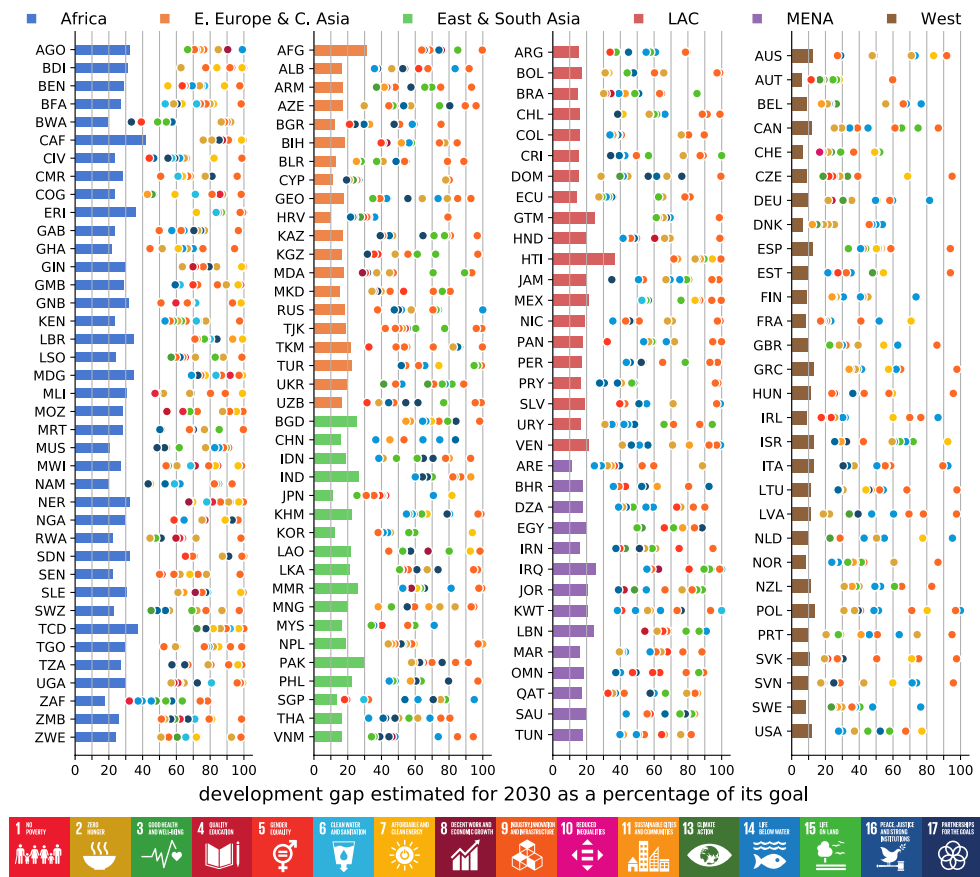
The reader should be aware that our methodology can deal with country-specific features such as the following: the network of interdependencies between indicators; the historical context reflected in the database and considered for calibration purposes; the indicators' initial conditions for prospective analyses; and the establishment of the 2030 goals attending to the countries' idiosyncrasies and political systems. Country-specific estimations are key when using the model for providing policy guidelines; however, technically this is not always possible with other methodologies. For example, in regression analyses using aggregate data, information from different countries has to be pooled

to obtain enough degrees of freedom. The latter approach precludes the possibility of making inferences for particular countries and, thus, the estimates have limitations in terms of policy advice.

### SDG gaps

We present our estimates of SDG gaps for 2030 at the level of each country in Figure 4. The bars indicate average levels across indicators, and the colored dots correspond to the ten indicators with the largest estimated gaps. The latter exemplify the gap disparities that exist within each country. As expected, the advanced market economies of the *West* exhibit gaps that are substantially lower than those estimated for the least developed countries (like those in *Africa*). However, there are also relatively successful countries in other regions of the world, such as Cyprus (CYP) and Croatia (HRV) in *E. Europe & C. Asia*; Japan (JPN), South Korea (KOR), and Singapore (SGP) in *East and South Asia*; and the United Arab Emirates (ARE) in *MENA*. In contrast, the least successful countries are Haiti (HTI) in *LAC*; and the Central African Republic (CAF), Eritrea (ERI), and Chad (TCD) in *Africa*.

At a more aggregate level, we can observe gap disparities across clusters and across SDGs. For example, while no country in *Africa* has an average gap below 18%, all



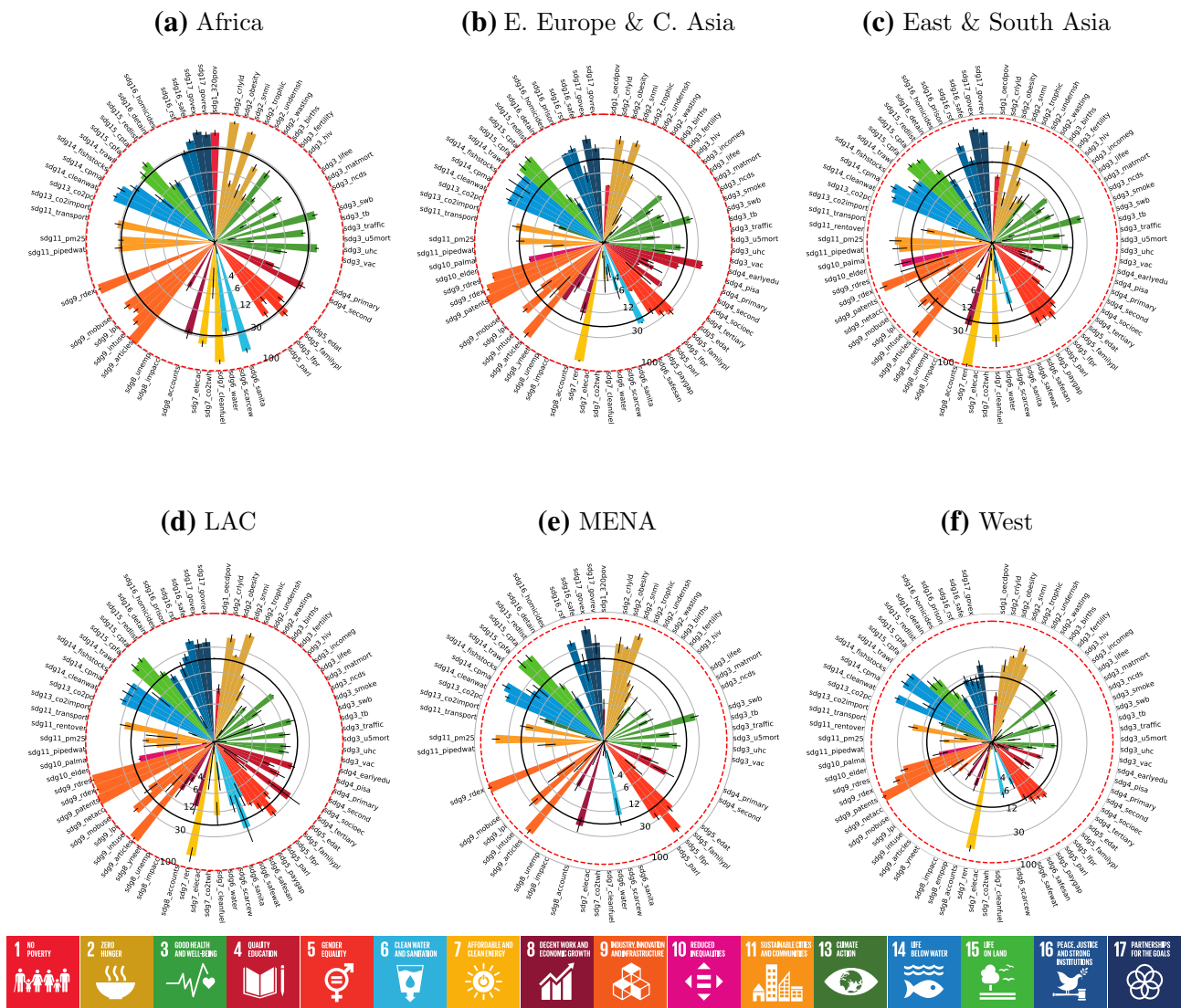
**Fig. 4** Average SDG gaps for 2030 by country. *Notes:* The bars denote the average SDG gap for 2030 (across all indicators of the country) for each individual country. Bars are colored according to the country clusters described in Figure 3. The dots correspond to the ten indicators with the largest estimated gaps. Each dot is colored

according to the corresponding SDG of its indicator. We use the model to estimate the indicators’ projections for 2030. For precise estimates and confidence intervals of each individual indicator gap, see Appendix C. *Sources:* Authors’ own calculations

the countries in the *West* have an average gap below 12%. The systematic persistence of certain dot colors (such as orange, corresponding to SDG 9 [Industry, innovation and infrastructure]) suggests that, in some SDGs, it is more difficult to close the gaps. Some of the most persistent SDGs across the dot markers are SDG 9—‘Industry, innovation and infrastructure’—and SDG 7—‘Affordable and clean energy’. Such a pattern is especially visible in *Africa*.

Figure 5 provides a complementary visualization of the SDG gaps. Here, the gaps of the indicators have been averaged across countries in the same cluster. These plots reveal that only one indicator of SDG 7 is persistently close to the 100% gap in *Africa*, and that several indicators of SDG 9

exhibit high gaps. Another feature revealed by this visualization is that most of the environmental indicators in SDGs 14 (Life below water) and 15 (Life on land) present gaps above the cluster average (identified with the solid black ring). The reader should be aware of the risks of aggregation, which are evident when comparing the gaps estimated for the indicators in SDG 2 (Zero hunger) in *Africa* and *West*. Here, *Africa* is expected to perform better than *West* in obesity, nitrogen emission, and human trophic levels (which relates to dietary diversity). These problems are endemic to advanced market economies, so our results are intuitive. However, if we were to aggregate these gaps for the whole SDG 2, the result would suggest a similar performance between both clusters



**Fig. 5** SDG gaps for 2030 aggregated by cluster and indicator. *Notes:* We use the model to estimate the indicators’ projections for 2030. The height of each bar represents the average gap between the SDG and the indicator level predicted by 2030 computed across countries in the cluster. Empty spaces between bars indicate that no data was available for the corresponding indicator in any country from the cluster. The solid black ring corresponds to the average gap across

countries (in the cluster) and indicators. The dashed red ring indicates the largest average gap (between indicators in the cluster). The black lines at the top of each bar denote the  $\pm$  standard error of the mean gaps across the countries of a cluster. For the full name of the indicators see Table B.1 in Appendix B. For precise estimates and confidence intervals of each individual indicator gap, see Appendix C. *Sources:* Authors’ own calculations

since the indicators related to hunger and malnutrition show the opposite performance (so their gaps would cancel out each other, at least approximately). Clearly, even with a multidimensional view of development, there exist specific policy issues that perform in substantially different ways across countries, even if they belong to the same dimension. This is one of the reasons why it is so important to move beyond the common practice of pooling cross-national and SDG-level data, and to produce more granular estimates that reflect the context and the spending capabilities of each country in each indicator.

### Sensitivity to changes in the budget size

A key issue to be addressed when studying the feasibility of SDGs is the impact that budgetary changes have on the evolution of social, economic, and environmental indicators. In the context of this paper, we are interested in understanding how sensitive are the different SDG gaps to changes in public expenditure. Thus, we estimate the country-specific sensitivity of each indicator to changes in the overall size of the budget during the 2020–30 period. Our dataset suggests substantial variation in the growth of public spending between the 2000–2010 and the 2010–2020 decades (an average of 47%). Thus, our estimates consider prospective simulations with positive and negative changes of up to 50% with respect to the historical expenditure levels reflected in the data.<sup>23</sup> We measure sensitivity by calculating the difference between gaps from a benchmark scenario that maintains the historical expenditure levels (the average yearly expenditure from the data, projected over 10 years) and a scenario that considers changes in the size of the budget.

Figure 6 presents a highly disaggregated picture of the different sensitivities when the budget is increased by 50%. Larger markers denote more sensitivity, while the gray lines indicate the absence of an indicator in a particular country. As a reference point, the largest marker corresponds to a reduction of 13% in an SDG gap. From this visualization we can highlight several important results. First, there is substantial heterogeneity across countries—positioned in the vertical axis—and indicators—positioned in the horizontal axis—with respect to the magnitude of gap reductions. Second, the most notorious impacts are not randomly scattered, but rather concentrated in specific SDGs (compare columns of different colors) and indicators. For instance, two gaps in SDG 9 (‘Logistics performance index’ and ‘Mobile broadband subscriptions’) have notable reductions in most

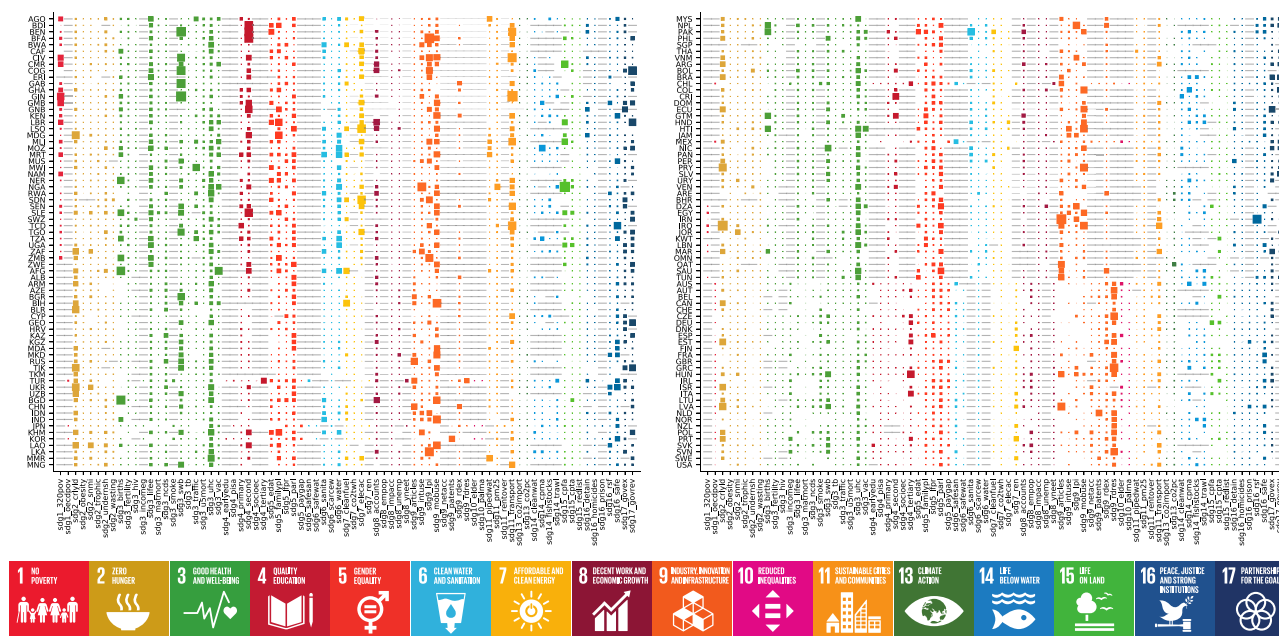
of the countries where data are available. Third, the gaps of economic indicators in SDG 8 (Decent work and economic growth) are not particularly sensitive to a 50% increase in the budget size, especially when compared with those of SDG 9 (see the size of brown markers *versus* that of orange markers). Fourth, with the exception of some African cases, the gaps in SDGs 13 (Climate action), 14, and 15 (the environmental ones) rarely exhibit substantial improvements. Fifth, excluding a few country-indicator cases, the SDG 16 (Peace, justice and strong institutions) gaps do not seem responsive to a 50% increase in the budget. In section 4.3, we show that these diverse sensitivities are the result of long-term structural factors that impose a constraint to the effectiveness of public expenditure in government programs. Before elaborating on these structural factors, we provide further sensitivity results related to reductions to the budget size, and to an alternative sensitivity metric.

Figure 7 presents sensitivity results for a 50% reduction in budget sizes. In this case, the sensitivity outcomes mean that the SDG gaps widen. As a reference point, the largest marker corresponds to a gap augmentation of nearly 20% with respect to the benchmark case. This implies that, in general, SDG gaps are more sensitive to a 50% reduction than to an increment of the same proportion in the budget size. This sensitivity asymmetry becomes evident when contrasting the outcomes of SDG 8—in ‘Adults with an account at a bank or other financial institution’—presented in Figures 7 and 6. A similar asymmetric pattern can be found in environmental indicators from SDGs 14 and 15.

To have a better understanding of the asymmetric sensitivity between a 50% increment and reduction in the budget, we would like to revisit three assumptions about our modeling approach. First, we aim to model short-term dynamics and, hence, long-term structural factors are given through the exogenous parameters  $\alpha_i$  that are specific to each indicator and country. Second, the impact of the public funds devoted to the different government programs is viewed in the context of short-/mid-term effects. This is so because we model a probability  $\gamma_{i,t}$  representing the chance of indicator  $i$  to improve in the subsequent period  $t + 1$ . These two aspects are combined into the evolution equation 1. While more public spending increases  $\gamma_{i,t}$ , the long-term structural factors  $\alpha_i$  limit the growth speed. Therefore, government expenditure only affects  $\gamma_{i,t}$ , not  $\alpha_i$ . Third, public spending is a necessary condition for development. Thus, from looking at the evolution equation we can tell that, if less expenditure brings  $\gamma_{i,t}$  close to zero, then the growth trials are almost always unsuccessful, so the indicator dynamics stagnate.

From the side of budgetary increments, there seems to be a limit to how much some gaps can be reduced in a given period while, on the side of reductions, no improvement can be expected in the absence of public funds. Furthermore,

<sup>23</sup> An expenditure growth scenario for the next 10 years may be hindered thanks to the COVID-19 global pandemic.



**Fig. 6** SDG gap shrinkage due to a 50% increment in per capita expenditure. **Notes:** The size of the markers is proportional to the reduction of the SDG gap caused by an increase in government

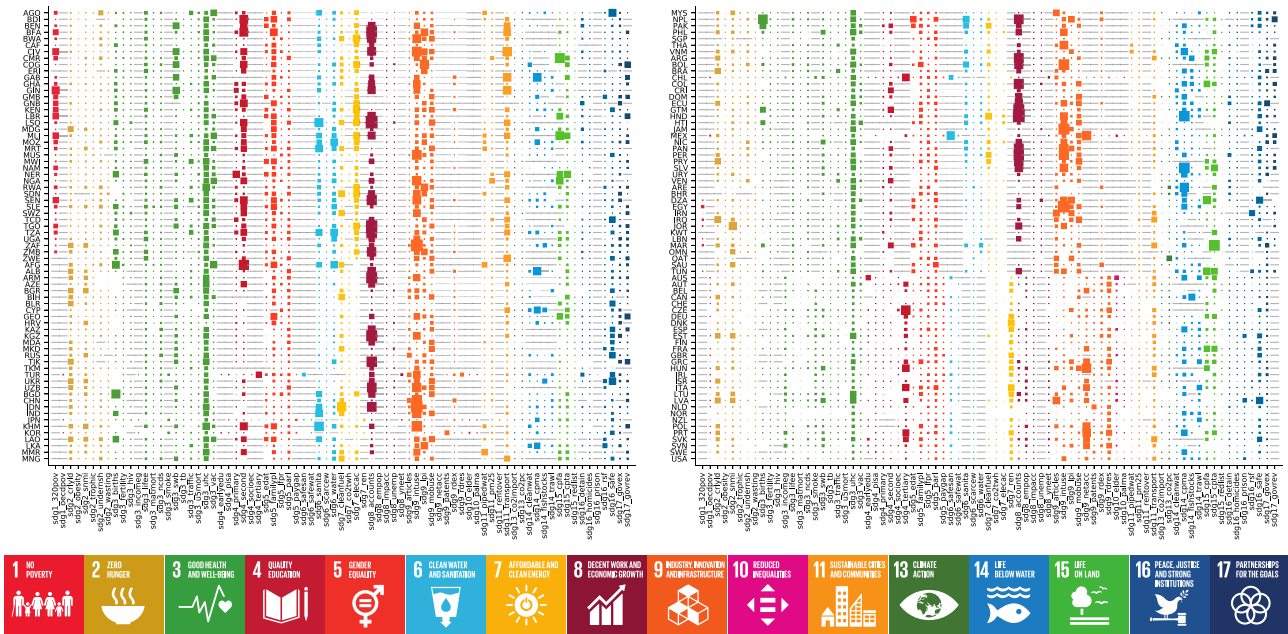
given the law of motion of the indicators, and other micro-foundations of the model, the response to expenditure changes may vary in non-linear ways. Thus, to provide a full picture of these non-linear response functions, we measure sensitivity in terms of the number of years that it would take to achieve a certain level (for each indicator and country), and compute them for 1% variations (positive and negative) in the budget size. We present the results aggregated into SDGs and clusters. In Figure 8, we present the aggregate response functions in the range of budgetary changes between -50% to +50% (with marginal changes of 1%).<sup>24</sup> We calculate the response functions using the difference in the number of years it takes for an adjusted budget to reach the levels of the indicators obtained in 2030 with the benchmark scenario. In the latter calculation, the historical annual expenditure average is projected forward throughout the following decades. If the budgetary changes produce additional years, there is a delay, while if there are some saved years, the delay displays in a negative scale.

spending. The biggest marker corresponds to the largest reduction in the sample. The gray lines indicate the absence of an indicator in a particular country. **Sources:** Authors' own calculations

Due to the aforementioned problems of aggregating indicators, the results presented in 8 should be considered as qualitative evidence of the non-linear responses to changes in public expenditure.<sup>25</sup> First, note that every SDG shows certain level of sensitivity to both positive and negative budgetary changes. Second, confirming our previous findings, the sensitivity to positive and negative budgetary changes are systematically asymmetrical in terms of the response magnitude. Third, the sensitivity rankings across SDGs vary between clusters and depends on the magnitude and direction of the budgetary change. For example, for countries in *West*, SDG 13 is the most sensitive to budgetary reductions, but the same ranking position is not observed in other clusters. Nevertheless, it is important to emphasize that SDG 13 systematically exhibits important delays in all clusters.

<sup>24</sup> The curves in Figure 8 are composed of indicators that were able to converge during a set of Monte Carlo simulations. Thus, because there are selection biases due to the exclusion of non-converging indicators, these curves should be considered a qualitative result about the non-linear nature of development outcomes to budgetary changes.

<sup>25</sup> We provide the data of the country-indicator specific responses in [http://github.com/ogurrer/sdg\\_feasibility](http://github.com/ogurrer/sdg_feasibility).



**Fig. 7** SDG gap growth due to a 50% reduction in per capita expenditure. **Notes:** The size of the markers is proportional to the increase of the SDG gap caused by a reduction in government spending. The big-

gest marker corresponds to the largest reduction in the sample. The gray lines indicate the absence of an indicator in a particular country. **Sources:** authors' own calculations

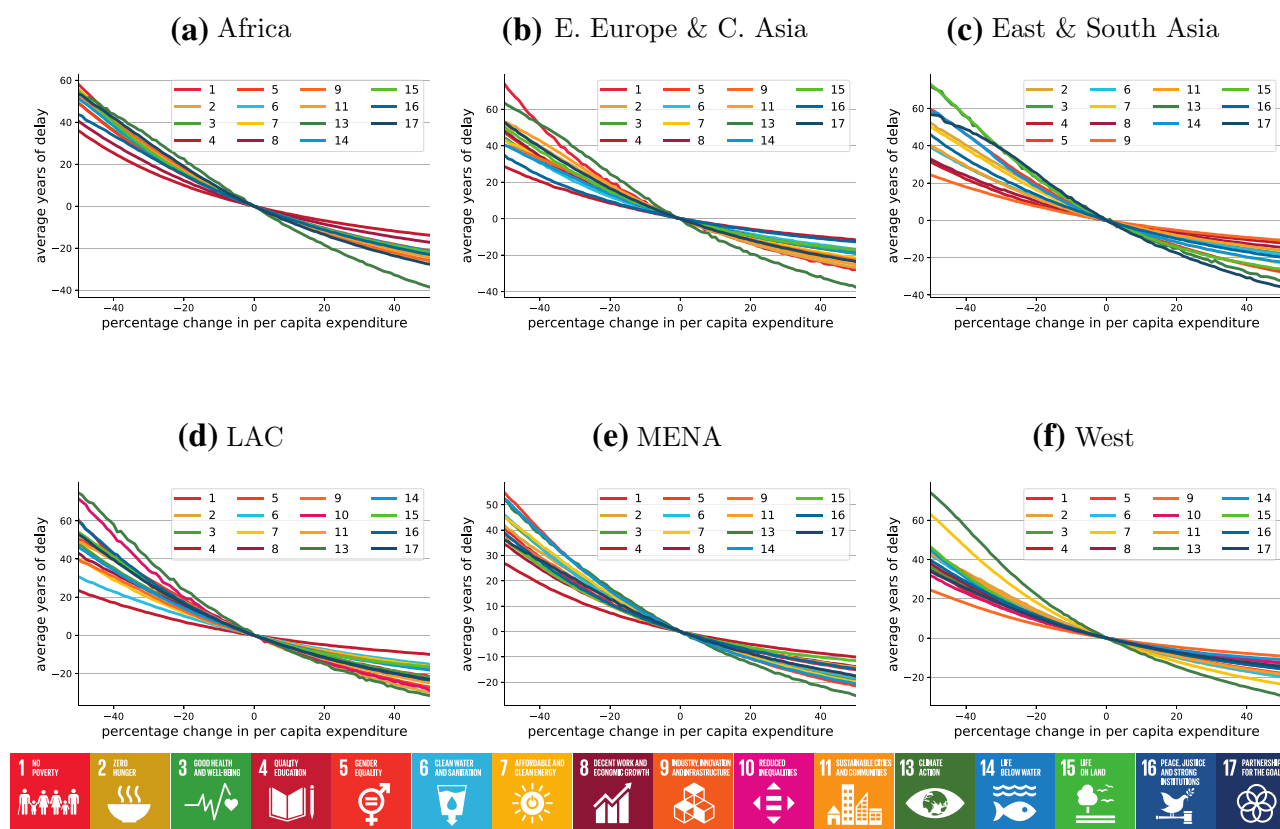
## Budgetary frontiers and structural bottlenecks

Now that we have established the existence of non-linear responses of the SDG gaps to public spending, we elaborate on their structural origins. Let us open our argument by stating the obvious: *that every government is constrained by time and resources*. Thus, in a short-/mid-term scenario, time is critical in order to achieve a set of goals. While a particular policy may succeed in improving an issue, the amount of improvement is constrained by factors such as infrastructure, organizational practices, individuals' incentives, and technology that can only be modified through changes to the existing government programs; changes that take place in a longer time span. Thus, in the scope of existing government programs, these factors are considered exogenous, and we capture them through parameter  $\alpha_i$ . It follows that the success in reaching development goals is partly determined by how much  $\alpha_i$  allows an indicator to improve during a set amount of time. Not knowing these limits to success could lead to ineffective policy priorities and bad planning in terms of long versus short-/mid-term policies.

To unearth the limits imposed by structural factors, it is useful to think about the following hypothetical question: *How much, in the years left to reach the SDGs, can the SDG*

*gaps be closed if public funding was unlimited and fully efficient?* This theoretical scenario removes the resource constraints from the equation, and leaves us with the interaction between structural factors and time. Therefore, by estimating the SDG gaps under this hypothetical setting, it is possible to establish bounds to how small an SDG gap can become by 2030. To achieve this, we only need to assume  $\xi(\gamma_i) = 1$  in equation 1. When  $\xi(\gamma_{i,t}) = 1$  for every indicator, we say that the country operates at the 'budgetary frontier'. Thus, the SDG gaps that remain open at this frontier describe the limitations of increasing expenditure in the current government programs. In other words, if an SDG gap remains open at the budgetary frontier, it means that—regardless of how much public expenditure increases—the strategy will be unsuccessful if the long-term structural factors (i.e., their bottlenecks) in that policy issue are not addressed.

Figure 9 presents the average SDG gaps at the budgetary frontiers of the different countries in the sample. Panel (a) aggregates the gaps across indicators with each country. Note that none of the average gaps closes entirely, even in the most advanced nations. As expected, these gaps are wider in *Africa*, reaching 39% in the Central African Republic (CAF). This diagram illustrates the relevance that local features have in the wide disparities observed across countries' performances. The estimated SDG gaps at the budgetary frontiers show countries exhibiting structural long-term hindrances of different magnitudes, even if they belong to the same cluster. Although the model cannot



**Fig. 8** Changes in convergence time as a function of the budget size. **Notes:** These response functions are calculated for each cluster (panel) and SDG (colored lines), averaging across indicators. The horizontal axis denotes the increment or reduction of the annual budget during the decades following 2020. A positive value in the vertical axis indicates the number of additional years that it would

take to reach the levels originally projected for 2030 (hence, when the budget change is zero, all the lines collapse at zero in the y-axis). A negative value in the y-axis translates into years saved to reach the 2030 levels. The reference levels are determined in the baseline scenario used for Section 4.1. The numbers in the inset labels indicate different SDGs. **Sources:** Authors' own calculations

distinguish the specific reasons behind these discrepancies, it is reasonable to argue that their causes lie in bottlenecks of a local nature.

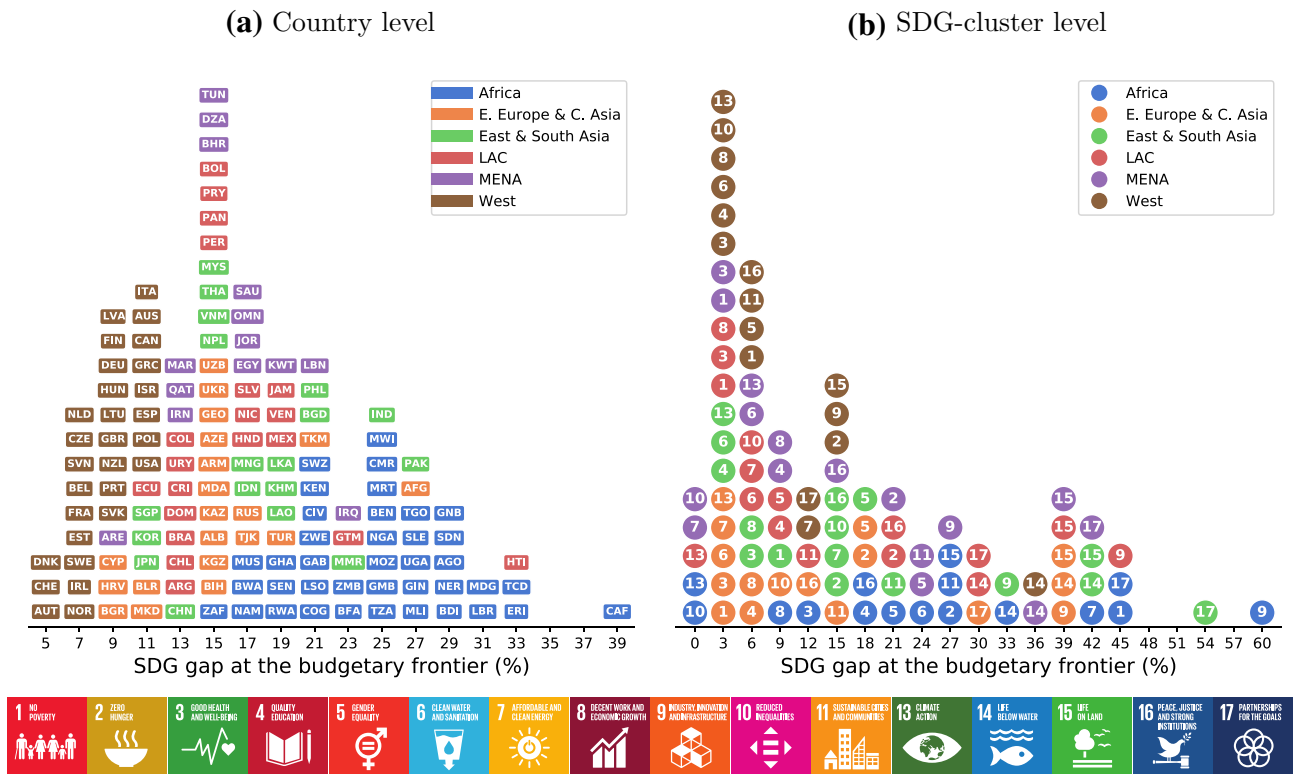
The right panel in Figure 9 shows the average gaps at the budgetary frontier, aggregated into SDGs within each cluster. The fact that SDG 13 presents a near-null gap in countries from *Africa* and *LAC* indicates that environmental issues related to climate action could be improved, on average, by properly funding existing government programs

in those regions. However, this is not the case for other environmental SDGs. For instance, in SDGs 14 and 15, the frontier gaps vary between 27 and 42%<sup>26</sup>.

Finally, a cautious reader may consider that public spending should have structural consequences, so the exogenous factors  $\alpha_i$  could also be affected in the short term. While this reasoning is, in principle, correct, the empirical evidence suggests that this process is rather weak. For instance, if the structural factors contained in  $\alpha_i$  were to change substantially in the short term, then the SDG gaps estimated from simulations using more recent data samples should

<sup>26</sup> SDG 15 for *West* is an outlier with a gap of 15%.





**Fig. 9** Budgetary frontiers. *Notes:* A country operating at the budgetary frontier has a  $\xi(\gamma_{i,t}) = 1$  for every indicator  $i$  and every period  $t$  (see equation 2) in the methods section). At the budgetary frontier, the only frictions slowing down the indicators’ growth are in the structural parameter  $\alpha_i$ . Panel (a): budgetary frontiers calculated by

averaging gaps across indicators for each individual country. Panel (b): budgetary frontiers calculated by averaging gaps across countries and indicators at the level of SDGs for each cluster. The average gaps have been discretized to produce the visualizations. *Sources:* Authors’ own calculations

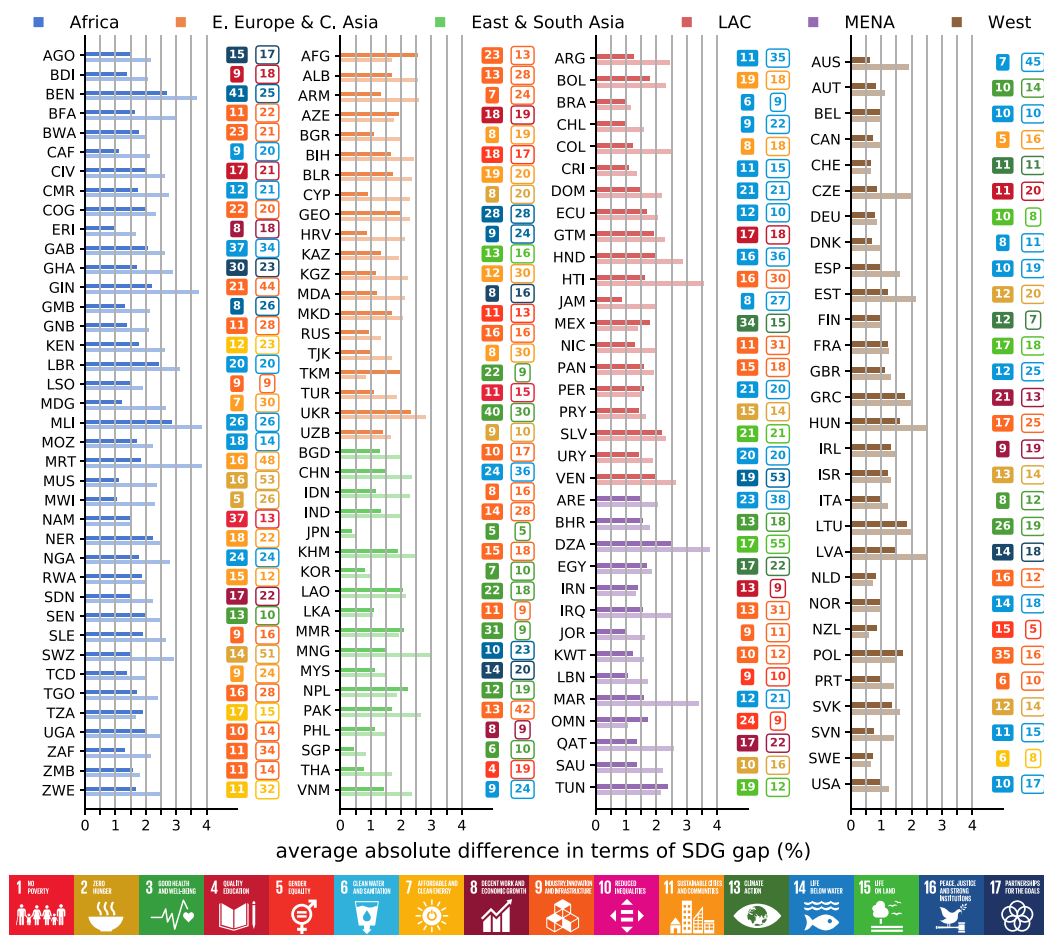
significantly differ. To demonstrate that this is not the case, we calibrate the model and perform the same analysis as in Sect. 4.1 but, instead of using the full 21-year dataset (with 2000–2020 coverage), we employ a 10-year (2011–2020) and a 5-year (2016–2020) sample.<sup>27</sup>

Figure 10 shows that our original estimates are robust to these alternative samples, as the six clusters show relatively small differences in their average gaps (see Appendix I for more disaggregated yet robust results).<sup>28</sup> For calculating

these differences, we compare the average SDG gap for each country produced in the benchmark simulations—using 21 years of data—and the gap estimated with smaller time series of historical data (either 10 or 5 years). Notice also that the closer the size of these time series is to the whole historical sample, the smaller the difference in the average gaps is. That is to say, the dark bars are smaller than the light ones. From this, we conclude that the SDG network and the structural factors exhibit slow dynamics, validating our

<sup>27</sup> This involves re-estimating the network, the structural parameters, and the gaps.

<sup>28</sup> These are differences in the average gaps. The numbers to the right of these bars show the most sensitive indicators to the length of the time series considered. The first column of numbers corresponds to comparisons with 10-year time series, while the second to comparisons with 5-year time series.



**Fig. 10** Robustness to different sampling lengths. *Notes:* The bars indicate the average absolute difference in estimated gaps (in percentage) between the benchmark case—using 21 years of data—and one where the model was calibrated with shorter time series. The dark bars are calculated using the model calibrated with 10-year time series. The light bars are computed using the model calibrated with

with 5-year time series. The solid squares on the right of each panel denote the color of the SDG to which the most sensitive indicator belongs in the case of differences using 10-year time series. The hollow ones correspond to 5-year time series. For a more disaggregated presentation of these results see Appendix I. *Sources:* Authors' own calculations

conceptualization of long- versus short-/mid-term effects. Accordingly, the budgetary frontiers involve long-term considerations and demand the implementation of innovative micro-policies.

## A discussion on the model's strengths and limitations

Models of multidimensional development typically use composite indices such as the Human Development Index and the SDG Index. However, if analysts wish to provide more nuanced advice with respect to specific SDGs in terms of policy prioritization and budgetary allocations, it is necessary to model the evolution of each separate dimension without aggregating them and losing valuable information. This task is problematic for statistical/econometric and machine

learning approaches since they cannot deal easily with a high-dimensional policy space characterized by few observations (short time series). For example, multi-output models (such as regressions of equation systems or neural networks) demand unrealistically large amounts of observations for each dimension/indicator. To overcome this limitation, analysts pool cross-national data to produce their estimates. This, however, has the costly implication of removing country-specificity because any interpretation from the estimated parameters is limited to a hypothetical country with the average characteristics of the sample. In addition, data-pooling strategies only work with a limited number of indicators, since there exist only so many countries.

In data-fitting approaches, the problem of few observations aggravates when considering interdependencies between indicators because the number of potential interactions (parameters to be estimated) grows exponentially

1. Much of the empirical quantitative literature—which policymakers often use to guide their decisions—focus on the impact of one (or multiple) indicator(s) on another (or others). However, indicators are not instruments that governments can directly manipulate, but rather endogenous variables resulting from spending decisions. Hence, governments often motivate their expenditure choices using studies that do not offer evidence on how effective or viable it would be to fund a particular SDG given the existing government programs. There are two alternatives to remedy this analytical hurdle: the use of granular expenditure data or the implementation of a generative model of public spending.
2. Highly granular data of public expenditure—properly linked to specific development indicators—are practically non-existent. Under these circumstances, analysts have to rely on data aggregated into a few broad sectors (e.g., education, health, poverty alleviation) and select a ‘representative’ indicator for each one—or an average index.
3. When constructing a dataset to use these methods, one must assemble a large cross-country panel to obtain the necessary degrees of freedom for making estimations possible.
4. Most data-fitting approaches can only consider one dependent variable, which is inconsistent with the systemic view of the SDGs.
5. Establishing a causal link between an expenditure variable and an aggregate indicator is problematic because of confounding factors and reverse causation (because a government can adjust its budget according to the observed performance of the indicators). This problem also applies to studies using more sophisticated machine learning methods.
6. Due to inefficiencies during the policymaking process and spillover effects, the level of expenditure in a policy issue does not reflect the actual amount of resources effectively used.

In general, data-fitting and aggregate models are ill-suited due to their lack of explicit causal mechanisms. To overcome this problem, computational approaches such as agent-computing models can be useful. Nonetheless, these models may also demand large amounts of data, so they are typically employed in micro-level studies relevant to a specific SDG. However, this does not mean that agent computing cannot be used for comprehensive analyses of SDGs but rather that substantial efforts need to be made in this direction. For instance, Allen et al. (2016) provide an extensive review of model types used for the assessment of SDGs. They find that agent-computing models account for only 1% of the studies in their literature survey.

In this paper, we contribute to this effort by developing an agent-computing model that is explicit about a critical causal channel: *public expenditure*. In contrast with recent studies on SDGs, like those mentioned in the paper’s introduction, our model can be calibrated for individual countries on a large policy space, helping researchers and practitioners to get the most out of the available data. Furthermore, it does not impose aggregate relationships between the different indicators. Rather, it is very flexible since it allows the user to introduce any network of interdependencies that are relevant to the context under study (Ospina-Forero et al. 2020). While our model is not explicit about the full complexity of the system (and no other model is), it provides a rich enough yet parsimonious specification, which facilitates counterfactual experiments (‘what if’ scenarios) and allows estimating the impact of budgetary changes.

The proposed computational method also has some limitations that the reader should be aware of. Our model cannot produce *ex ante* evaluations of new government programs, nor can it yield policy prescriptions if, in the out-of-the-sample analysis, there is a drastic transformation in the technological, political, and organizational underpinnings of a country. In this sense, the model assumes a ‘business as usual’ setting in which the system keeps working with similar government programs and structural features as those prevalent during the sampling period. This assumption is realistic in short-term analyses (less than 6 years) and admissible for evaluating policy design in a medium-term setting (5–15 years). Thus, our approach focuses on short-/mid-term effects, and it explicitly separates structural factors that shape long-term dynamics. Ironically, while much of the existing methods suffer from the same limitations, their applications often tend to emphasize long-term scenarios.

The robustness of the model can be enhanced when disaggregated expenditure data are available at the SDG or government program levels. However, these databases only exist for a very few countries. Therefore, in this paper, we offer a worldwide application of the model and show that it can provide insightful policy guidelines even if a country only has aggregate expenditure data. As in any quantitative approach, when more detailed empirical information is available, our model can generate more specific policy prescriptions. For instance, with expenditure data disaggregated at the SDG level, it is possible to establish whether different budgetary allocations, to those observed historically, could exert an impact on the closing of development gaps.

## Conclusion

We propose a bottom-up computational framework to analyze the short- and mid-term impact of budgetary allocations in a large set of SDG indicators. Our simulations use data

from individual countries. Hence, it allows specifying context-dependent settings: initial conditions, calibrated parameters, and spillover effects among indicators. The underlying theory assumes fixed structural factors in the indicators' evolution equations and exogenous network topologies (which could be constructed in tandem with other qualitative and quantitative approaches). Our approach is useful to understand how to allocate resources across the existing government programs, so it facilitates identifying key priority areas. Moreover, through counter-factual simulations, the model can discover bottlenecks associated with the inefficiency of the public expenditure, which is key to achieve any development agenda. However, the model is not designed to identify the causes behind the structural constraints that prevent a country from closing its SDG gaps. Hence, the outputs are not informative about how to reformulate the existing micro-policies or how to generate new ones.

Our main results provide novel and nuanced estimates of the development gaps that will remain open in 2030, at the level of each country and indicator. We also find that more government spending is not enough to close the SDGs gaps, even if countries were operating at a budgetary frontier that entails enough resources for the existing government programs. Hence, complementary micro-policies are ultimately needed to overcome structural—long-term—bottlenecks and to improve the relevant indicators. When looking at the model's estimates, we can offer detailed interpretations of the simulation results. For instance, some environmental concerns such as clean air can be substantially ameliorated with a larger budget, while others (e.g., SDGs 14 and 15) require undertaking well-designed government programs to shift the historical course of ineffective policies.

Despite the simplicity of the model, it is possible to use it to infer two crucial country-specific features: (1) the possibility of closing the SDG gaps, and (2) the existence of long-term bottlenecks. Therefore, when analysts can identify one or several government programs with a particular indicator, it is possible to establish some policy guidelines with the model's estimates. Depending on the values of these features and the level of the indicator's historical performance, it is possible to define different routes of policy action. That is to say, whether the program should be reviewed, in terms of incentives and organizational practices before spending more public funds, or whether the SDG gaps can be closed by just channeling more funds into the existing programs.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11625-022-01095-1>.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes

were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Akenroye T, Nygård H, Eyo A (2018) Towards implementation of sustainable development goals (SDG) in developing nations: a useful funding framework. *Int Area Stud Rev* 21(1):3–8
- Allen C, Metternicht G, Wiedmann T (2016) National pathways to the sustainable development goals (SDGs): a comparative review of scenario modelling tools. *Environ Sci Policy* 66:199–207
- Amos R, Lydgate E (2020) Trade, transboundary impacts and the implementation of SDG 12. *Sustain Sci* 15(6):1699–1710
- Aragam B, Gu J, Zhou Q (2019) Learning large-scale bayesian networks with the Sparsebn package. *J Stat Softw* 91:11
- Asadikia A, Rajabifard A, Kalantari M (2021) Systematic prioritisation of SDGs: machine learning approach. *World Dev* 140:105269
- Benedek D, Gemayel E, Senhadji A, Tieman A (2021) A post-pandemic assessment of the sustainable development goals. *Staff Discussion Notes*, 2021(003)
- Boeren E (2019) Understanding sustainable development goal (SDG) 4 on “quality education” from micro, meso and macro perspectives. *Int Rev Educ* 65(2):277–294
- Carley K (1996) Validating computational models. Working Paper. CASOS Program, Pittsburgh, PA
- Castañeda G, Chávez-Juárez F, Guerrero O (2018) How do governments determine policy priorities? Studying development strategies through networked spillovers. *J Econ Behav Organ* 154:335–361
- Castañeda G, Guerrero O (2018) The resilience of public policies in economic development. *Complexity*
- Castañeda G, Guerrero O (2019a) The importance of social and government learning in ex ante policy evaluation. *J Policy Model*
- Castañeda G, Guerrero O (2019) Inferencia de Prioridades de Política para el Desarrollo Sostenible. Reporte Metodológico, Programa de las Naciones Unidas para el Desarrollo
- Castañeda G, Guerrero O (2019) Inferencia de Prioridades de Política para el Desarrollo Sostenible: El Caso Sub-Nacional de México. Reporte Técnico, Programa de las Naciones Unidas para el Desarrollo
- Castañeda G, Guerrero O (2019) Inferencia de Prioridades de Política para el Desarrollo Sostenible: Una Aplicación para el Caso de México. Reporte Técnico, Programa de las Naciones Unidas para el Desarrollo
- Collste D, Pedercini M, Cornell SE (2017) Policy coherence to achieve the SDGs: using integrated simulation models to assess effective policies. *Sustain Sci* 12(6):921–931
- de Sousa J, Mayer T, Zignago S (2012) Market access in global and regional trade. *Reg Sci Urban Econ* 42(6):1037–1052
- de Wolff T, Cuevas A, Tobar F (2021) MOGPTK: the multi-output gaussian process toolkit. *Neurocomputing* 424:49–53
- Dhami S (2016) *The foundations of behavioral economic analysis*. Oxford University Press, Oxford
- Dhaoui I (2018) Achieving sustainable development goals in MENA countries: an analytical and econometric approach

- Efron B (1981) Censored data and the bootstrap. *J Am Stat Assoc* 76(374):312–319
- Fader M, Cranmer C, Lawford R, Engel-Cox J (2018) Toward an understanding of synergies and trade-offs between water, energy, and food SDG targets. *Front Environ Sci* 2:2
- Fuso Nerini F, Sovacool B, Hughes N, Cozzi L, Cosgrave E, Howells M, Tavoni M, Tomei J, Zerriffi H, Milligan B (2019) Connecting climate action with other sustainable development goals. *Nat Sustain* 2(8):674–680
- Gaulier G, Zignago S (2010) BACI: International trade database at the product-level. Technical Report 2010-23, CEPII
- Gobierno del Estado de México (2020). Informe de Ejecución del Plan de Desarrollo del Estado de México 2017-2023; a 3 Años de la Administración
- González-Pier E, Barraza-Lloréns M, Beyeler N, Jamison D, Knaut F, Lozano R, Yamey G, Sepúlveda J (2016) Mexico's path towards the sustainable development goal for health: an assessment of the feasibility of reducing premature mortality by 40% by 2030. *Lancet.Global health*, 4(10):e714–e725
- Guerrero O, Castañeda G (2020a) Policy priority inference: a computational framework to analyze the allocation of resources for the sustainable development goals. *Data & Policy*, 2
- Guerrero O, Castañeda G (2020) Quantifying the coherence of development policy priorities. *Dev Policy Rev* 00:1–26
- Guerrero O, Castañeda G (2021) Does expenditure in public governance guarantee less corruption? Non-linearities and complementarities of the rule of law. *Econ Gov* 22(2): 139–64
- Guerrero O, Castañeda G, Trujillo G, Hackett L, Chávez-Juárez F (2021) Subnational sustainable development: the role of vertical intergovernmental transfers in reaching multidimensional goals. *Socio-Econ Plan Sci*, p 101155
- Ionescu G, Firoiu D, Tănăsie A, Sorin T, Pîrvu R, Manta A (2020) Assessing the achievement of the SDG targets for health and well-being at EU level by 2030. *Sustainability* 12(14):5829
- Izquierdo A, Pessino C, Vuletin G, editors (2018) Better spending for better lives: How Latin America and the Caribbean can do more with less. Inter-American Development Bank
- Jones B, Baumgartner F, Breunig C, Wlezienc C, Soroka S, Foucault M, François A, Green-Pedersen C, Koski C, John P, Mortensen P, Varone F, Walgrave S (2009) A general empirical law of public budgets: a comparative analysis. *Am J Polit Sci* 53(4):855–873
- Kroll C, Warchold A, Pradhan P (2019) Sustainable development goals (SDGs): are we successful in turning trade-offs into synergies? *Palgrave Commun* 5(1):1–11
- Luken R, Mörec U, Meinert T (2020) Data quality and feasibility issues with industry-related sustainable development goal targets for Sub-Saharan African countries. *Sustain Dev* 28(1):91–100
- Lusseau D, Mancini F (2019) Income-based variation in sustainable development goal interaction networks. *Nat Sustain* 2(3):242–247
- Machingura F, Lally S (2017) The sustainable development goals and their trade-offs. Technical report, Overseas Development Institute, London, United Kingdom
- McGowan P, Stewart G, Long G, Grainger M (2019) An imperfect vision of indivisibility in the sustainable development goals. *Nat Sustain* 2(1):43–45
- Mensi A, Udenigwe C (2021) Emerging and practical food innovations for achieving the sustainable development goals (SDG) target 2.2. *Trends Food Sci Technol* 111:783–789
- Moyer J, Hedden S (2020) Are we on the right path to achieve the sustainable development goals? *World Dev* 127:104749
- OECD (2019) Governance as an SDG accelerator: country experiences and tools. OECD Publishing
- OECD (2020) Measuring the distance to the SDGs in regions and cities
- Ospina-Forero L, Castañeda Ramos G, Guerrero O (2020) Estimating networks of sustainable development goals. *information and management*
- Osuji E, Nwani S (2020) Achieving sustainable development goals: does government expenditure framework matter? *Int J Manag Econ Soc Sci (IJMESS)* 9(3):131–160
- Pedercini M, Arquitt S, Chan D (2020) Integrated simulation for the 2030 agenda. *Syst Dyn Rev* 36(3):333–357
- Pedercini M, Arquitt S, Collste D, Herren H (2019) Harvesting synergy from sustainable development goal interactions. *Proc Natl Acad Sci* 116(46):23021–23028
- Philippidis G, Shutes L, M'Barek R, Ronzon T, Tabeau A, van Meijl H (2020) Snakes and ladders: world development pathways' synergies and trade-offs through the lens of the sustainable development goals. *J Clean Prod* 267:122147
- Porciello J, Ivanina M, Islam M, Einarson S, Hirsh H (2020) Accelerating evidence-informed decision-making for the sustainable development goals using machine learning. *Nat Mach Intell* 2(10):559–565
- Pradhan P, Costa L, Rybski D, Lucht W, Kropp J (2017) A systematic study of sustainable development goal (SDG) interactions. *Earth's Future* 5(11):1169–1179
- Pradhan P, Subedi D, Khatiwada D, Joshi K, Kafle S, Chhetri R, Dhakal S, Gautam A, Khatiwada P, Mainaly J, Onta S, Pandey V, Parajuly K, Pokharel S, Satyal P, Singh D, Talchabhadel R, Tha R, Thapa B, Adhikari K, Adhikari S, Bastakoti R, Bhandari P, Bharati S, Bhusal Y, Bk B, Bogati R, Kafle S, Khadka M, Khatiwada N, Lal A, Neupane D, Neupane K, Ojha R, Regmi N, Rupakheti M, Sapkota A, Sapkota R, Sharma M, Shrestha G, Shrestha I, Shrestha K, Tandukar S, Upadhyaya S, Kropp J, Bhuju D (2021) The COVID-19 pandemic not only poses challenges, but also opens opportunities for sustainable transformation. *Earth's Fut* 9(7):e2021EF001996
- Putra M, Pradhan P, Kropp J (2020) A systematic analysis of water-energy-food security nexus: a south Asian case study. *Sci Total Environ* 728:138451
- Sachs J, Schmidt-Traub G, Kroll C, Lafortune G, Fuller G, Woelm F (2020) Sustainable development report 2020. Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN), New York
- Sobczak E, Bartniczak B, Raszkowski A (2021) Implementation of the no poverty sustainable development goal (SDG) in visegrad group (V4). *Sustainability* 13(3):1030
- Sulmont A, de Alba García, Rivas M, Visser S (2021) Policy priority inference for sustainable development: a tool for identifying global interlinkages and supporting evidence-based decision Making. In *Understanding the Spillovers and Transboundary Impacts of Public Policies*, OECD Publishing, Paris
- United Nations (2020). SDG indicators, United Nations Global SDG Database
- Warchold A, Pradhan P, Kropp J (2021) Variations in sustainable development goal interactions: population, regional, and income disaggregation. *Sustain Dev* 29(2):285–299
- World Bank (2017) World development report 2017: governance and the law. International Bank for Reconstruction and Development / The World Bank, Washington
- World Bank (2020). SDG atlas 2020
- Zelinka D, Amadei B (2019) A systems approach for modeling interactions among the sustainable development goals part 2: system dynamics. *Int J Syst Dyn Appl (IJSDA)* 8(1):41–59