

Assessing qualitative and quantitative dimensions of uncertainty in energy modelling for policy support in the United Kingdom

Steve Pye^{a,e*}, s.pye@ucl.ac.uk, Francis G. N. Li^a, Arthur Petersen^b, Oliver Broad^c, Will McDowall^c, James Price^a, and Will Usher^d

^a UCL Energy Institute, Central House, 14 Upper Woburn Place, London, WC1H 0NN, United Kingdom

^b UCL Department of Science, Technology, Engineering and Public Policy (STePP), Boston House, 36-37 Fitzroy Square, London, W1T 6EY, United Kingdom

^c UCL Energy Institute, Central House, 14 Upper Woburn Place, London, WC1H 0NN, United Kingdom

^d Environmental Change Institute (ECI), University of Oxford

^e MaREI Centre, Environmental Research Institute, University College Cork, Cork, Ireland

* Corresponding author

Highlights

- NUSAP clarifies the strengths and weaknesses of the modelling knowledge base
- Highly influential model assumptions on CCS and bioenergy shown to have a weak pedigree
- The approach employs broad peer engagement, which engenders understanding and critique
- Recommendation that such an approach is embedded in analytical processes supporting policy

Abstract

Strategic planning for the low carbon energy transition is characterised by a high degree of uncertainty across many knowledge domains and by the high stakes involved in making decisions. Energy models can be used to assist decision makers in making robust choices that reflect the concerns of many interested stakeholders. Quantitative model insights alone, however, are insufficient as some dimensions of uncertainty can only be assessed via qualitative approaches. This includes the strength of the knowledge base underlying the models, and the biases and value-ladenness brought into the process based on the modelling choices made by users. To address this deficit in current modelling approaches in the UK context, we use the NUSAP (Numeral Unit Spread Assessment Pedigree) approach to qualify uncertainty in the energy system model, ESME. We find that a range of critical model assumptions that are highly influential on quantitative model results have weaknesses, or low pedigree scores, in aspects of the knowledge base that underpins them, and are subject to potential value-ladenness. In the case of the UK, this includes assumptions around CCS deployment and bioenergy resources, both of which are highly influential in driving model outcomes. These insights are not only crucial for improving the use of models in policy-making and providing a more comprehensive understanding of uncertainty in models, but also help to contextualise quantitative results, and identify priority future research areas for improving the knowledge base used in modelling. The NUSAP approach also promotes engagement across a broader set of stakeholders in the analytical process, and opens model assumptions up to closer scrutiny, thereby contributing to transparency.

Keywords

Uncertainty analysis; NUSAP; qualitative dimension of uncertainty; decarbonisation; energy systems modelling

1.0 Introduction

1.1 Energy and climate strategy under uncertainty

Strategic planning for the low carbon energy transition is characterised by a high degree of uncertainty across many knowledge domains and by the high stakes involved in making decisions. The future availability and costs of transition technologies, the political environment under which they may be deployed, and the role of changing societal preferences and individual behaviours are key uncertainties for decision makers to contend with, and which will impact numerous stakeholders [1]. The UK has long identified the need for decarbonisation of the energy system, with legally binding national targets on emissions reduction that remain among the most ambitious globally [2]. A well-developed science policy architecture has developed over the last decade to explore and implement the transition [3,4], and some progress has been made in recent years, particularly in the power generation sector [5]. However, strategic decisions in a number of critical sectors have yet to be taken, for example in areas such as switching away from natural gas for heating in buildings, the decarbonisation of freight transportation, and how to address the growing emissions from aviation. The long-lived nature of critical infrastructures for supplying energy makes path dependencies and lock-in to legacy assets a real risk [6].

This type of challenge, where urgent near-term choices must be made in an environment where perfect information and universal agreement amongst key stakeholders is impossible to achieve, is characterised in the scientific literature as the domain of *post-normal science* [7,8]. This is in direct contrast to the definition of *normal science* by Kuhn [9], where observations are used to iteratively resolve testable hypotheses through experimentation. The assessment of strategic options in a *post-normal science* context, such as long term energy policy, must contend with multiple epistemic uncertainties that arise from our imperfect knowledge, including those that can be quantified in modelling tools, but also those that are not easily quantifiable.

Van der Sluijs [10] argues that most quantitative-only approaches do not adequately deal with those dimensions of uncertainty that are non-quantifiable. These include the strength of the underlying knowledge base, the level of theoretical understanding of the processes modelled, and the value-ladenness coproduced by modellers themselves because of the requirement to make choices across key model assumptions. As an illustration, a quantitative analysis performed for a particular policy problem might produce modelling results which suggest that a given input parameter is highly influential on the distribution of costs of meeting a given objective. But what is typically missing from such an exercise is an assessment of the uncertainty arising from the strength of the knowledge base underpinning that quantified model outcome. Such non-quantifiable uncertainty, were it exposed to decision-makers, might reduce the perceived robustness of the model-derived quantitative insight, and lead to different conclusions for policy.

Approaches that recognise this multi-dimensional nature of uncertainty, as described in the next section, can provide decision makers with a more comprehensive understanding of uncertainty and improve the robustness of the resulting choices made. They help avoid quantitative-only approaches which only consider a “*restricted agenda of defined uncertainties – ones that are tractable*” [11]. When faced with policy challenges in the *post-normal* domain, a broad approach to uncertainty assessment is vital. It is entirely possible that “*unquantifiable uncertainties dominate the quantifiable ones*” [10] and excluding them from the analysis will risk giving decision makers a highly restricted perspective on the range of possible outcomes. The challenge is that uncertainties are numerous and appear, as per Walker’s typology [12], at different stages across the modelling process, from the problem framing itself, the selection of model input parameters, the structural design and process of defining relationships, and from the subjectivity of model users.

1.2 Existing approaches and knowledge gaps

Energy models are likely to continue to play a key role in ongoing energy transitions by providing the evidence base for planning policies on climate mitigation [13], which in turn serve as key drivers behind many transitions towards sustainability [14]. The Paris Agreement [15] recommends mid-century low emission strategies and states that individual signatories must provide regular updates on their strategic plans for low carbon development (Nationally Determined Contributions, NDCs), forcing a requirement for policymakers to assess low carbon energy transitions at the country-scale [16]. Energy models provide a clear framework for systematic experiments that explore the possible consequences of the multiple different options in systems that are otherwise difficult to grapple with [17]. In the UK, for the last 15 years, energy models have been critical in providing guidance on system decarbonisation, notably in relation to the affordability and feasibility of achieving targets, on issues of path dependency, and on identifying critical feedbacks and linkages between sectors [3,4]. However, the treatment of uncertainties in strategic analysis has had a number of limitations, and practitioners in the modelling community have repeated calls for increased use and improvement of methods for uncertainty analysis [18–20]. This is mirrored by calls from strategy analysts, industry experts and government decision makers, who are cognisant of a broad spectrum of future uncertainties facing the energy transition and also the limitations of current modelling and scenario analysis practices to capture them [1,21].

Energy systems analysis in the UK (and in many other national and regional contexts) has typically focused on the use of scenarios for exploring different futures [22–25]. Historically it has been common for analysts to employ a handful of scenarios with deterministic inputs, using a coupled storyline-and-simulation approach [26]. This approach has been shown to have limitations – ex post analysis of modelled energy futures based on scenario analyses often finds that real world developments occur that are completely outside of the anticipated range [27,28]. Modelling practitioners are increasingly drawing from a range of more advanced quantitative techniques to assess parametric uncertainties as a means of capturing more of the problem space in their work. Techniques found in the UK context include probabilistic analysis [19,20], stochastic programming [18], and modelling-to-generate-alternatives (MGA) [29,30] using parameter uncertainty ranges supported through expert elicitation [1,31,32].

While valuable for opening up dialogue and highlighting the uncertain nature of the knowledge claims made in this field, none of the above techniques alone are able to adequately identify and assess those non-quantifiable dimensions of uncertainty discussed earlier. An innovative approach to assessing uncertainties in model-based analysis is the NUSAP system [10]. NUSAP, or Numerical Unit Spread Assessment Pedigree, was first proposed by Funtowicz and Ravetz [7], before undergoing substantial development and implementation in the Dutch Government’s applied policy research institutes [33]. NUSAP retains the strengths of quantitative uncertainty assessment but brings a focus on the qualitative assessment of the quality or ‘pedigree’ of the underlying model assumptions. This framework, which includes both standard uncertainty analysis techniques but also assessment of non-quantifiable uncertainties, increases the robustness of emerging conclusions from models, providing decision makers with an enhanced understanding of the strengths and weaknesses of model insights.

1.3 Aims and objectives of the paper

In this paper, we demonstrate how practitioners can broaden the scope of strategic advice given to energy system decision makers by holistically considering both qualitative and quantitative dimensions of uncertainty using the NUSAP approach. For this research, we have used a prominent UK energy systems model, the Energy Systems Modelling Environment (ESME) [34], which is under active development and has been used for academic [35,36], industry [37] and government [38] applications. We describe the application of the NUSAP protocol for assessing the qualitative dimension of uncertainty and show how this can be combined with insights from a quantitative mathematical sensitivity analysis (using the Morris Method).

The NUSAP system has been used before in diverse scientific fields such as the assessment of acid rain and transboundary air-pollution impacts, the global integrated assessment of climate policies, and the effects on human health of waste disposal practices [10,39–41]. The life cycle assessment community have also effectively used pedigree scoring of underlying data assumptions, which is a key element of the NUSAP approach, to better recognise its impact on uncertainty [42,43]. This is the first time such an approach has been applied to a national energy model used to inform thinking on energy transitions towards deep decarbonisation. Additionally, we incorporate novel elements into the NUSAP approach, such as the assessment of model pedigree in multiple time horizons.

The key research questions for this study were as follows:

- i. What are the key non-quantifiable uncertainties arising from limitations in the knowledge base underlying the ESME model?
- ii. How do they inform and complement our understanding of uncertainty from quantitative uncertainty approaches, and what are the implications for strategic energy transition planning, in terms of policymaking and future research needs?

The paper is structured as follows; in Section 2, we describe the approach taken to the uncertainty analysis. This includes descriptions of the NUSAP-based workshop used to elicit non-quantifiable uncertainties and of the Morris Method global sensitivity analysis that was used to explore the quantitative uncertainties. Section 3 presents the results from the workshop, while Section 4 provides a discussion of the results. The paper concludes with recommendations for modelling uncertainty, particularly in support of decision-making, and a discussion on future research needs.

2.0 Research design and methods

In this section, we first describe the NUSAP approach, followed by a brief overview of the model ESME, the subject of the approach. We then describe how we selected the model assumptions to focus on, and the process of designing and implementing a NUSAP workshop, to elicit an understanding of uncertainties in ESME.

2.1 NUSAP

The NUSAP approach¹ was proposed in 1990 [7] and further developed to meet the need to revise the approach to uncertainty assessment following strong criticism of the Netherlands Environmental Assessment Agency, MNP,² in the late 1990s, on the credibility and reliability of the assumption and models used in environmental policy [33,44]. At the core of the approach is the recognition that uncertainty is multi-dimensional, with dimensions including technical (inexactness), methodological (unreliability), epistemological (ignorance) and societal (social robustness) [10]. Quantitative uncertainty assessments, particularly those using global sensitivity analysis, typically representing the *technical* dimension, provide information on the uncertain quantifiable assumptions that have the greatest influence on the results [45].

However, missing from such assessments is information on the quality of the underlying knowledge base that allows users of the model results to form a view on the robustness of the emerging conclusions. Applying the NUSAP approach allows for scrutiny of the ‘pedigree’ or quality of model assumptions by a range of experts. Pedigree is judged based on scoring against multiple criteria, capturing different aspects of (non-quantifiable) uncertainty (Table 2, Section 2.3). The term ‘model

¹ NUSAP stands for Numeral-Unit-Spread-Assessment-Pedigree. ‘Numeral’ is the numbers used in the analysis; Unit is the units for the numbers; ‘Spread’ is the variance in the model outputs due to input uncertainty; ‘Assessment’ represents qualitative judgements about assumptions; and ‘Pedigree’ represents the strength of knowledge and use of knowledge in the analytical approach.

² MNP was merged with Netherlands Institute for Spatial Research (RPB) in 2008 to form the PBL Netherlands Environmental Assessment Agency (*Planbureau voor de Leefomgeving*)

assumption’ relates to the following aspects of the model where uncertainty could manifest itself, and follows the typology proposed in Walker et al. [12] and subsequent PBL guidance [46]; i) model structure (relationships embedded in equations), and ii) model inputs, such as data, boundary conditions, and analysis framing.

For each model assumption considered, experts score against the selected criteria, as described in section 2.3. The resulting pedigree scores provide insights into qualitative dimensions of uncertainty, which can then be used to compliment the quantitative uncertainty analysis. A key strength of the approach is that both components can be brought together using a diagnostic diagram, as shown in Figure 1 [47]. The pedigree scores highlight the ‘strength’ of the model assumption, reflecting the methodological and epistemological limitations of the knowledge base, while the quantitative sensitivity analysis shows the influence of the model assumptions on the results via statistical methods.

The diagnostic diagram can be separated into four quadrants. In Q1, the influence of assumptions on model outputs is low, and the underlying knowledge base is strong, resulting in limited cause for concern. The same may also be true for Q3, where despite a weaker knowledge base, the influence of assumptions on outputs is low. In Q2, the assumptions have a strong effect on the overall modelled outcome, but the strong pedigree of the knowledge base can give decision makers confidence that the model is providing useful information to inform decisions about the real world system. In Q4, assumptions have high statistical influence over model outcomes, but a weak pedigree. This quadrant can be characterised as the ‘danger zone’. In this region, what appears to be an important outcome from a quantitative perspective may need to be treated with caution as the quality of the knowledge underpinning the assumption has been revealed to be weak or contested. The application of this diagnostic diagram is further considered in the results section (Section 3.3).

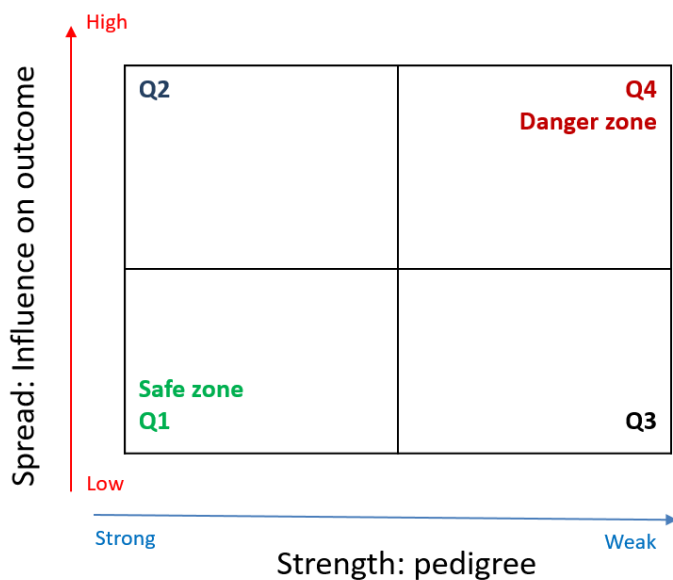


Figure 1. Diagnostic diagram used to combine quantitative and qualitative dimensions of uncertainty

Another important feature of the NUSAP approach is that it allows for in-depth scrutiny of model assumptions through expert stakeholder involvement, highlights the strengths and weaknesses of the model analysis (and areas to prioritise for improvement), and also explicitly clarifies the distinct areas of disagreement across assumptions. In effect, it provides greater transparency to the analytical process, for example through identifying a diversity of views on the strength of the underlying knowledge base, and the level of consensus across different assumptions. The evaluation of previous NUSAP exercises by participants has consistently highlighted broad agreement amongst stakeholders regarding their value [39,48]. However, the NUSAP process of eliciting expert

perspectives via a workshop is time intensive and therefore not necessarily always appropriate in a fast-moving policy context. Klopprogge et al. [40] suggest therefore that it may be most appropriate to apply the NUSAP approach in high value analysis cases – for example, where decision stakes are high and uncertainty can have a large impact on policy development, and also in cases where societal controversies are prevalent. Hence, analysis for climate and energy policy is a good example of where such an approach can be highly relevant.

2.2 The ESME model

In this paper, we apply the NUSAP approach to ESME³, a regionally disaggregated model of the UK energy system, used to determine the role of different low carbon technologies for achieving the mid- to long-term climate mitigation goals set in UK legislation. The model has been used extensively for informing government and industry strategies, and underpins a range of research papers on different aspects of energy system decarbonisation [19,20,35,49]. Built in the AIMMS environment, the model uses linear programming to assess cost-optimal technology portfolios. It covers the key sectors of the UK energy system, including power generation, industry, buildings, transport and other conversion sectors e.g. biofuel production, hydrogen production.

A set of energy service demands that capture requirements for industrial and building heating, electricity supply and mobility needs are provided as exogenous inputs, and largely reflect Government projections. The model then endogenously determines how to meet these demands in a cost-optimal manner, through investment in end use technologies (including energy efficiency measures), and in production and supply options to provide for different energy forms. Primary resource supply is characterised by commodity price and resource availability, with no distinction between imports and domestic indigenous production (except for biomass), and no explicit representation of resource and upstream sectors (although these are accounted for implicitly through the energy balances, prices and other statistics used as inputs). For emissions accounting, the model accounts for CO₂ but not other greenhouse gases (GHGs), such as methane (CH₄) and nitrous oxide (N₂O). Therefore, the CO₂ emissions constraints applied in the model that are used to understand the implications of the UK's GHG emission reduction plans make an exogenous assumption about the level of non-CO₂ GHG levels in future years, taking account of expected abatement, and are adjusted accordingly. In this version of the model, a total constraint of 160 MtCO₂e is assumed, based on an assumed non-CO₂ GHG level of 55 MtCO₂e [50], allowing for CO₂ emissions of 105 Mt in 2050.

The uncertainties around the future costs and performance of different technologies and resource prices are captured via a probabilistic approach, where multiple scenarios are generated to explore a wide range of pathways. A full list of the assumptions considered in this NUSAP exercise are provided in Table 1, with further background information on each in Appendix B. A full description of the model is provided in Heaton [34], while the data assumptions and sources are also published [51].

2.3 Selecting the model assumption for assessment

The large number of model assumptions, the structural complexity of the ESME model (which contains numerous equations), and the in-depth nature of the NUSAP assessment process meant that careful consideration was needed regarding which assumptions to prioritise during the workshop in order to keep the process manageable. The guiding principle was to focus our analysis on the most influential model assumptions. Following the practice observed in earlier NUSAP workshops [39], a dual-layer approach was used for their selection; firstly, the model inputs which were ranked the highest in terms of their influence on model results (expressed in terms of the costs of meeting decarbonisation targets) were selected, based on two previously published global sensitivity analyses using ESME [20,52]. Such approaches allow for an understanding of 'how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input' [53]. One such approach is the Morris Method [54], also known as the Elementary Effects Method, which provides a

³ Assumptions reviewed were in v4.1

computationally efficient means of screening a model in order to determine the set of input parameters to which the model solution is the most sensitive. A wide range of input parameters are first characterised as uncertain, and then subjected to the Morris Method sampling approach that captures the uncertainty space across the parameter value ranges. The ESME model is then run multiple times, in each case to meet the UK’s climate legislation across the carbon budget periods and 2050 target. The sensitivity metrics are then calculated based on their influence on total system costs. This sensitivity analysis approach detailed in Usher [52] was used here to inform the uncertainty analysis presented in the results section of this paper, the approach for which is described in detail in Appendix E.

Secondly, an interview was held with expert modellers from the Energy Technologies Institute (ETI), the developers of ESME, to elicit their views on what they considered to be the critical assumptions affecting the model output. Prior to the interview, participants were provided with an overview of the NUSAP approach, and the purpose of identifying key uncertainties for discussion at a workshop. The interview was exploratory in style [1], with participants free to discuss any issues relevant to model uncertainty. This gave a broader perspective based on the developers’ experience of using the model for different applications, and led to the inclusion in the workshop assessment of specific model features associated with analysis framing and model structure, which can also be considered within the NUSAP framework [55].

Table 1 lists the model assumptions selected. Five were chosen on the basis of the global sensitivity analysis (GSA) and four from expert elicitation (EE) via the interviews with model experts, including two structural assumptions. The end result was a set of critical assumptions for modelling decarbonisation pathways, covering a range of different aspects of the modelling.

Table 1. Selection of assumptions for pedigree assessment

Assumption category	Assumption code	Assumption	Uncertainty location	Method of determination*
Resources	BioRES	<i>Bioenergy resource potential</i> : limits the level of domestic biomass available for use in the energy system	Input data	GSA
	BioEF	<i>Bioenergy emission factors</i> : reflecting assumptions around carbon neutrality, impacting on the attractiveness of bioenergy as a zero or low carbon source of energy	Input data / boundary conditions	EE
	GasPRC	<i>Gas resource costs</i> : based on UK Government projections reflecting international market prices	Input data	GSA
Technologies	CCSmbr	<i>Combined Cycle Gas Turbines (CCGT) w/ CCS build rate</i> : expressed as a constraint in the model as maximum capacity deployment per annum	Input data	GSA
	CCScap	<i>Combined Cycle Gas Turbines (CCGT) w/ CCS Capex</i> : expressed as expenditure requirement to install a unit of capacity	Input data	GSA
	NUCcap	<i>Nuclear Gen. III** Capex</i> : expressed as expenditure requirement to install a unit of capacity	Input data	GSA
Emissions accounting	NonCO2	<i>Non-CO₂ GHG emissions</i> (implicit in CO ₂ trajectory): level of emissions that has an impact on the CO ₂ target trajectory	Input data / boundary conditions	EE
Model structure	TchLRN	<i>Technology learning rates</i> : approach to incorporating projections in the model of the technology costs reductions over time, due to learning effects	Model structure	EE
	PerFOR	<i>Perfect foresight optimisation</i> : model formulation where all economic ‘decisions’ are made with a full knowledge of information relating to the future.	Model structure	EE

* GSA = global sensitivity analysis; EE = expert elicitation. ** Nuclear Gen III is a shorthand reference for third generation nuclear reactors, intended to represent the latest class of nuclear build designs, e.g. Areva's EPR, Westinghouse's AP1000

2.4 NUSAP workshop for eliciting information on the qualitative dimension of uncertainty

A NUSAP workshop was held on 14th September 2017 to explore the qualitative dimensions of uncertainty across a range of assumptions used in the ESME model. In this section, we describe the selection of experts for the workshop, the approach to appraising the assumptions, and the process of running the workshop.

Expert selection

19 experts attended the NUSAP workshop, from a range of disciplinary backgrounds and professions, and all with different perspectives on modelling, including data providers, developers, and users of outputs. Just over half of the participants were from academia (across eight universities), with the other half equally split between government, industry, and other research organisations and institutes. All participants were primarily UK-based, although a number of the academic group had overseas affiliations.

Approximately 30 experts were originally contacted to be involved in the workshop. Determining invitations was based on the requirement for the workshop to include a mix of three different types of stakeholders: i) recognised subject experts on the model assumptions of focus; ii) experienced users of energy models who interact with the decision making community; and iii) users of energy model outputs in the decision making community. All three groups bring different perspectives on the use of models, and the knowledge base underpinning the different analyses undertaken.

Similar to other approaches to expert selection found in energy policy research (e.g. [56,57]), our list of experts was based on purposive selection, identifying the relevant expert community. The resulting 19 experts who attended provided a broad range of expertise and a diverse range of perspectives on possible energy system developments, based on their disciplinary backgrounds and professional experience. Prior to the workshop, all participants were contacted and provided with an advance information pack that contained a description of the NUSAP concept, the list of assumptions for assessment, and the pedigree criteria and scoring framework. This was particularly to allow for familiarisation of the criteria definitions.

Criteria selection for pedigree assessment

Table 2 lists the criteria used to assess the pedigree of the model assumptions in Table 1. In broad category terms, the NUSAP assessment focuses on the methodological, epistemic, and societal dimensions to model pedigree. These criteria have been used in many previous NUSAP assessments; in particular, we draw from Van der Sluijs et al. [10,39] on methodological and epistemological dimensions, and Kloprogge et al. [40] and Van der Sluijs [58] on the societal dimension. The methodological dimension focuses on uncertainty related to the perceived reliability of the estimates made for model inputs. The criteria for appraising reliability (in view of model purpose) considers how good a proxy a given assumption is, its empirical grounding, and whether it has been, or even if it can be, validated. The epistemological dimension focuses on areas of ignorance, in this case whether or not we think we have a sufficient and adequate theoretical understanding of the process to allow us to make meaningful assumptions about it.

The societal dimension recognises that the choice of model inputs is typically subjective, due to the different disciplinary backgrounds of model users, their political leanings, their preferences, whether there is limited information to inform assumptions, or whether there is wide range of possible choices that could be considered plausible. This *value-ladenness* is essentially unavoidable in sufficiently complex problems. Following the *post-normal science* tradition (see Section 2.1), NUSAP does not seek to reduce or minimise subjectivity, but instead considers that if there is a

plurality of viewpoints and perspectives then they must be recognised, and properly recorded as part of the assessment, to better understand the bearing that this might have on the conclusions that can be drawn.

Five criteria focus on the methodological and epistemological underpinning of modelling assumptions. These include: the extent to which the model is a good *proxy* for the real world system, the quality of the *empirical basis* for assigning values to model parameters, the perceived *rigour* of the methods used to derive input values for models from available empirical data, the extent to which input *validation* has been carried out, and the extent to which the *theoretical understanding* of the real world process is viewed as providing a reliable basis for estimates. A further three criteria focus on the societal robustness of the assumptions made. These include the *choice space* for making assumptions (i.e. how different could plausible input values be from one another) the *justification* for the choice of value used and how defensible this is perceived to be by others and whether or not there is likely to be *agreement amongst peers* in choosing model parameter values.

Table 2. Pedigree criteria used in NUSAP assessment

Uncertainty dimension	Criteria	Description
Methodological	Proxy	The extent to which the assumptions that we use in the model are proxies for the reality that we seek to represent, given the purpose of the model. Examples include over simplifications, first order approximations, incompleteness.
	Empirical Basis	The degree to which observations, measurements and statistics are used to estimate a parameter.
	Rigour	Refers to the norms for methodological rigour in this process applied by peers. Well-established and respected methods for measuring and processing the data would score high on this metric, while untested or unreliable methods would tend to score lower.
	Validation	The extent to which assumptions have been cross-checked and validated against other observations and measurements
Epistemological	Theoretical Understanding	The extent to which our theoretical understanding of the real world processes provides a reliable basis for estimates
Societal	Choice Space	The degree to which alternative choices of assumptions could be made i.e. the degree to which other acceptable / plausible assumptions are available
	Justification	The degree to which the approximation made in the model can be justified as a reasonable, plausible or acceptable assumption, given one's understanding of the reality. Can these assumptions be defended?
	Agreement Amongst Peers	The degree to which the assumption made in the model (by the analyst) is likely to coincide with other experts in the field

Development of scorecards

To guide the expert assessment of pedigree and provide a common platform for discussion and intercomparison, a structured scoring system was provided (as per NUSAP practice). This was based on successful systems used in past NUSAP assessments but was specifically tailored to the featured model and the model parameters chosen for the assessment. Scores between 0 to 4, where 0 denotes a weak pedigree and 4 a strong one, could be given for each criterion discussed in Table 2, and applied to each model assumption listed in Table 1. For each score, a description was provided to guide the experts in making judgments on assumption pedigree. A full scoring matrix is provided in Appendix A.

Our research for this paper added a novel component to the traditional NUSAP analysis process. Past assessments have typically explored the pedigree of models that explore research questions relating

to the present day or the near future. However, the ESME model is designed to assess energy and climate policy pathways across a very long time horizon from 2010 to 2050, and clearly many of the assessment criteria outlined above in Table 2 can be interpreted differently when considering the recent past as compared to the far future. In our pedigree assessment therefore, we introduced a separate scoring for those assumptions relating to the model base year (2010) and those relating to the long term future (2050). In the case of CCS technologies, which are not commercially deployed at scale and accordingly do not have 2010 base year assumptions, the years 2030 and 2050 were considered. It was important to recognise the considerable differences in the underlying knowledge base in these near-to-mid term (2010/2030) and longer term time periods, and therefore the associated level of uncertainty.

The scorecard, for use by the experts, included a range of information for assisting in scoring, including a description of each of the model assumptions, how it was incorporated into the ESME model, and the mathematical ranking of the input assumptions in terms of how much they affect model outputs (where relevant) from previous global sensitivity analyses. The full set of scorecard templates used are provided in Appendix B. A description of how the workshop was structured and how the scoring was undertaken is also provided in Appendix B.

3.0 Results

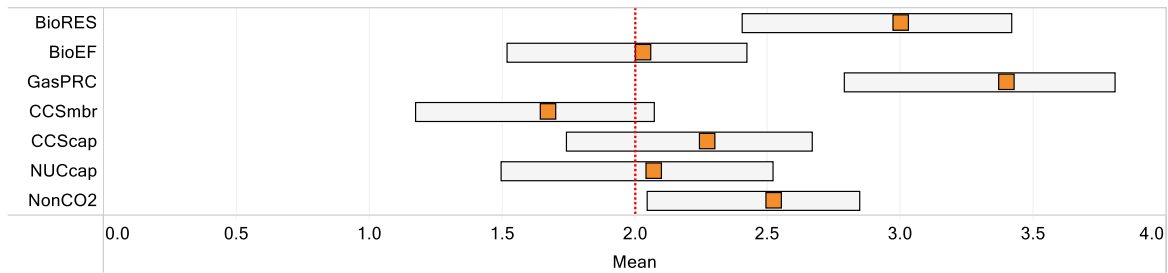
Selected results from the NUSAP workshop are presented in this section. We first note some general observations from the aggregated results before a discussion of the individual criteria scores across the groups of assumptions, including resources, technologies, emissions accounting and model structure. We conclude by describing a diagnostic diagram, which brings results from the workshop together with results from quantitative uncertainty analysis. Further detailed results, including the scoring by all participants for all of the individual assumptions from the workshop, are provided in Appendices C and D.

3.1 Aggregate pedigree scores

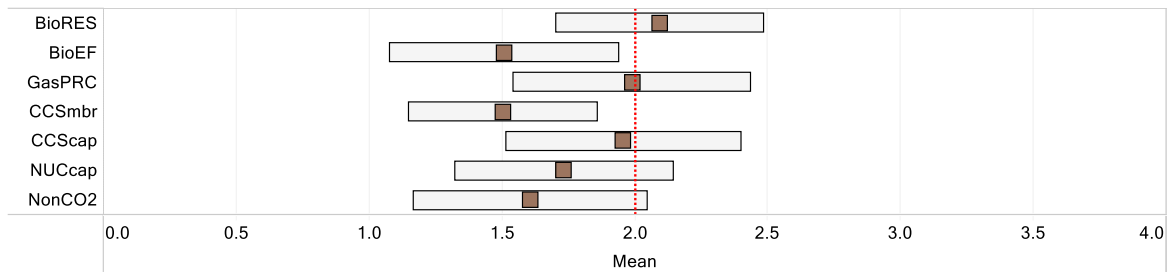
Aggregate scores for the pedigree of assumptions were calculated based on averaging the five methodological and epistemological criteria from each expert. These scores were then averaged for each of the assumptions, to provide aggregate pedigree scores, as shown in Figure 2. Standard deviations provide an indication of divergence in score between experts. The three value-ladenness criteria focus on the nature, not the strength, of the assumptions and are therefore not included in the aggregation. A clear difference emerges in scores between time periods, with near-term scores always being higher than those for 2050, reflecting the lower levels of uncertainty across the knowledge base for the current energy system. This distinction is particularly strong for bioenergy resource (BioRES) and gas prices (GasPRC), where average scores in 2010 are typically around 3.0 and 3.5 respectively, compared to the ranges 1.0-2.0 and 1.5-2.0 in 2050. Recognising that pedigree decreases in future years is important when considering the robustness of model insights in the long term.

For current or near term assumptions (

a) 2010 (2030 for CCS assumptions)



b) 2050



c) Structural

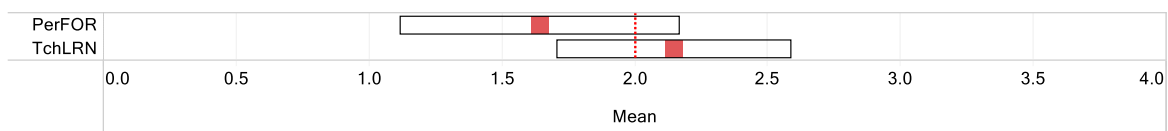


Figure Figure 2a), high scores for BioRES and GasPRC are observed. Lower scores are found for bioenergy emission factors (BioEF), nuclear capital expenditure (NUCcap), and build rates of gas CCS plant (CCSmbr). For BioEF, this reflects uncertainty in our knowledge of emissions from a very diverse resource base, and the highly aggregated proxy representation of the resource in the model. The ESME model only distinguishes between average emission factors for domestic and imported biomass, whereas in reality, almost every single bioenergy resource stream (based on type, location, processing, transport, end use technology) will have different emission factors. Given the diversity of emission factors found across the resource base, this assumption was given a relatively low *proxy* score, while the possible range of plausible values (*choice space*) was viewed as large (section 3.2).

A lower score for NUCcap reflects the limited evidence on the costs of contemporary nuclear power plant designs, particularly given the small number of existing projects from which to draw from. This uncertainty was further reflected in the low scores for the choice space criteria (many respondents suggested that the plausible range for nuclear capital costs is wide) and the agreement amongst peers criteria (respondents suggested that there is only limited agreement amongst experts). Across all assumptions, CCSmbr scored lowest, perhaps predictably as an assumption relating to the year 2030. Additional reasons for the low pedigree score for CCSmbr include the limited evidence base from which to draw from, and the fact that the non-cost barriers that this constraint includes are poorly understood, such as the societal acceptability of large-scale CCS, and the political will behind any successful future deployment. Interestingly, workshop participants rated the model assumptions about the capital cost level for the same technology (CCScap) as scoring higher due to empirical evidence today on the costs of different components of the CCS system, based on existing projects, and a view amongst the experts that a stronger knowledge base underpins this assumption.

The average scores for all model assumptions being assessed in 2050, shown in Figure 2b are lower and much more clustered together than their equivalent scores for 2010, reflecting the increasing uncertainty in the longer term. The lack of spread between 2050 average scores (in the range of 1.4 to 2.2), and similar score variation between assumptions sees most experts suggesting a much

weaker pedigree. For example, BioRES drops from a score of 3 in 2010 to 2 in 2050, reflecting expert views that future bioenergy resources are likely to be subject to multiple complex factors that are highly uncertain, around land use planning, competition between energy and food crops, and competition between the production of biomass for energy and the other non-energy uses of biomass (such as wood used in manufacturing). This weaker epistemic status of future values is not surprising; the implications of this for how decision makers think about model outputs is an important issue to further consider. For the CCS assumptions, the difference between the 2030 and 2050 scores is not marked, as the understanding of key parameters relating to this technology are uncertain at both points in time due to its low technology readiness level and the lack of real world examples where it has been deployed.

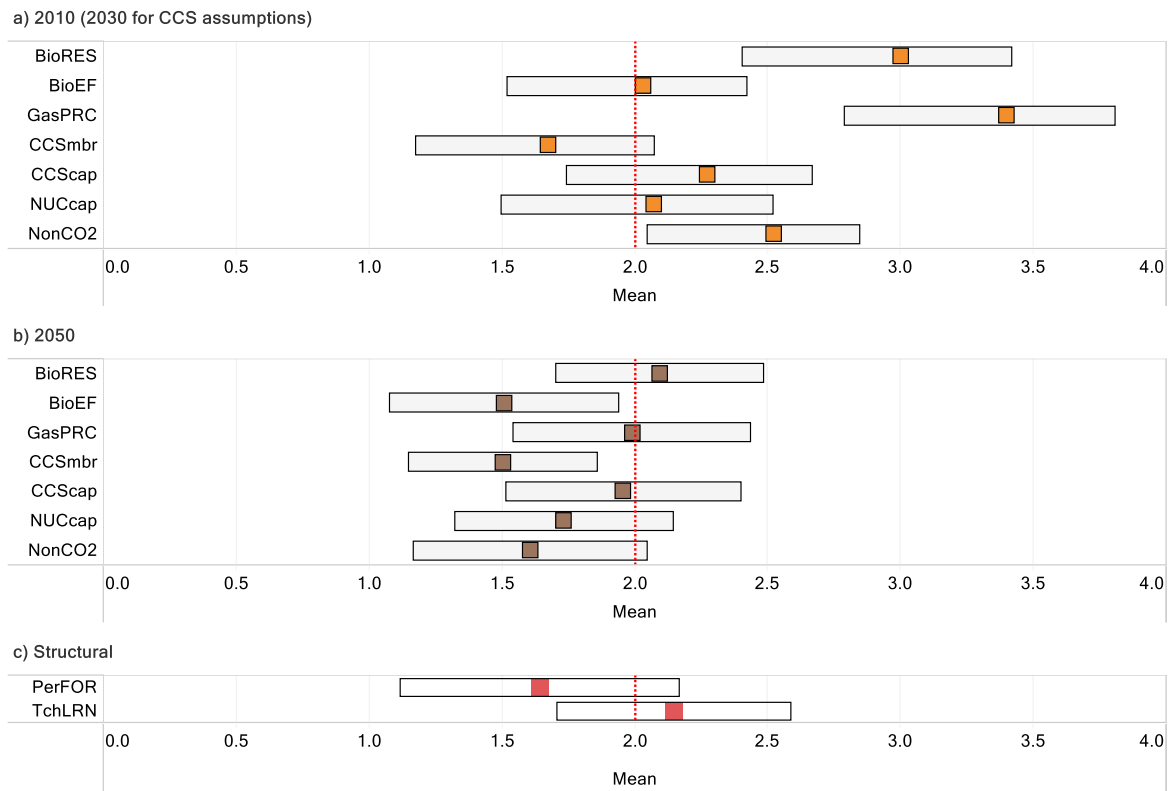


Figure 2. Mean aggregated pedigree scores. a) presents 2010 scores, except for the two CCS assumptions which relate to 2030. b) presents 2050 scores, and c) structural assumptions, technology learning and perfect foresight, which are both time independent. These pedigree scores show the mean and standard deviation of the participants mean scores based on the five pedigree criteria. The red line provides a reference point for comparison, as the central score value of 2. Model assumption labels are described in Table 1.

All assumptions see their relative scoring positions maintained between 2010 and 2050 with the exception of GasPRC and Non-CO₂. The level of non-CO₂ GHG emissions in the model is important because it determines the stringency of the cap on CO₂ emissions, which is the target that the ESME model tries to work towards (as covered earlier in Section 2.2). These future levels were considered much more uncertain by participants due to the many different factors influencing these emissions, from policy action, environmental factors such as land use management, and lifestyle choices around diets. On gas prices, the lower pedigree score (relative to other assumptions) in 2050 is related to multiple factors. These include uncertainty as to the how global supply and demand dynamics for natural gas play-out in future, given geo-political uncertainties, and the future potential role for gas in a future decarbonised energy system, which may be largely determined by the availability and deployment of CCS technologies.

Two structural assumptions are also shown in Figure 2b. The modelling approach to technology learning (TchLRN), has the highest average score of any 2050 assumption. “Technology learning” is

a term used to refer to how models capture future changes to technology costs resulting from factors like better and cheaper manufacturing processes (improvements over time). In the ESME model this is handled as an exogenous parameter i.e. the future cost pathways for key technologies are input assumptions that are decided in advance of the model being run. More complex approaches that link the change in technology costs to the level of technology deployment in the model are also possible. Experts' comments reveal a sense of pragmatism here given the model purpose, scale and practices across the modelling community; ESME is a model of national scale, and the UK typically has not been observed to drive learning across energy technologies in recent years. Furthermore, the practicalities of endogenising domestic learning effects is technically challenging, with limited data to parameterise such an approach.

Conversely, the other structural assumption, perfect foresight (PerFOR), gains a lower score. "Perfect foresight" is a structural feature of the model whereby the pathway solutions generated are made by taking into account the various changes to technology performance and costs in all future time periods. A number of experts did not view this as a structural assumption to be replaced by an alternative formulation but rather a core part of the model framework. There was also concern as to why this was being subject to appraisal as perfect foresight was not intended to be representative of how real systems work given the behaviour of different actors, but is rather a basis for understanding the role of different technologies from a techno-economic perspective. With differences of opinion as to how this should be scored, a number of experts (5) decided not to, and therefore the distribution of pedigree scores shown for this assumption has a smaller number of data points than the other scores.

3.2 Pedigree scores by criteria and assumption

The aggregate pedigree scores are useful for providing a summary but mask the wide variation in how individual assumptions were scored against each criterion, as shown in Figure 3 to Figure 5. For example, these figures show the extent to which participants thought that the representation of bioenergy in the model (as a relatively small number of homogenous feedstocks) is a good *proxy* for reality, and the quality of the *empirical basis* used to generate estimates of that bioenergy resource. Appendix D provides an alternative set of graphics, where the criteria for each model assumption are presented, instead of all assumptions for each criterion.

Of all the criteria, *validation* scores lowest in 2050 (range 1.0-1.5), reflecting that long-term model assumptions cannot be checked against real-world systems (Figure 3). For nearer term estimates, it is particularly useful to assess how well the modelled system is sense checked against current real-world system. For example, the *nuclear Capex* assumption (NUCcap) is the only assumption to have a score of less than 2.0, highlighting a potential weakness in validation against current nuclear projects, most of which are experiencing delays and incurring large cost-overruns. The *empirical basis* criterion, the degree to which observations, measurements and statistics are used to estimate a parameter has a similar pattern of scores to validation, albeit higher. Again, higher scores in 2050 are not obtainable as assumptions cannot be based on observations or measurements, but rather rely on estimates from other models (in a score range of 1.5-2.0). In 2010, scores are in the main between 2.5 and 3.5, where 3.0 represents a mix of observations and model-based estimates. As per the *validation* criterion, *empirical basis* scores for NUCcap and BioEF are lower, and are found closer to the central score of 2.0.

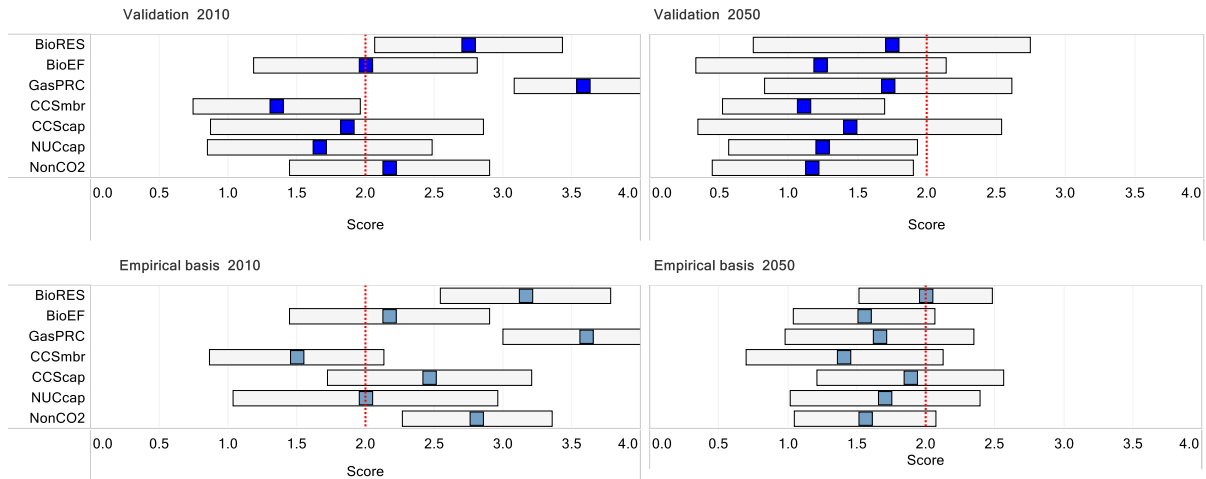


Figure 3. Mean and standard deviation for scores relating to validation and empirical basis criteria. The graphs on the left of the panel are for nearer term assumptions (2010, except 2030 for CCS) and on the right, for 2050. The red line provides a reference point for comparison, indicating the central score value of 2. Descriptions of the mode assumption abbreviations on the vertical axis can be found in Table 1. Scores across the structural assumptions can be found in Annex D (Figure D3).

Methodological rigour and *theoretical understanding* (typically in the range of 1.5-2.5) in 2050 tend to score better than *validation* and *empirical basis* (Figure 4). These scores reflect a stronger pedigree based on the use of rigorous methods for deriving assumptions, and a reasonable level of understanding of how the real world works that allows these assumptions to be determined. The relative scores across assumptions reflect the aggregate pedigree scores in Figure 2. For 2010, only two assumptions (GasPRC / BioRES) with a score of over 3.0 were deemed to have ‘good theoretical understanding’ and a ‘reliable method’. This highlights that for other assumptions, greater efforts may be needed to increase the pedigree with which the model represents base year conditions. NUCcap scores higher on *theoretical understanding* relative to other criteria, highlighting that for some assumptions, while a technology may be well understood, corresponding model estimates may still be poorly validated and have limited empirical grounding (as discussed above).

For the *proxy* criterion, or model representation of a given assumption, the scores in 2050 are clustered around a score of 2.0, described as a *moderate or acceptable representation*. It is notable that the two bioenergy assumptions score relatively lower against other assumptions under this criterion, reflecting the concern previously highlighted that the resource representation is too aggregated (as a single resource type) in the model. Given that the representation in the model of the different assumptions is the same in 2010 as it is in 2050, the divergence in scores is surprising, but perhaps reflects experts’ view of the *proxy* criterion as relating to both model representation in terms of methodological approach and in terms of data availability.

Value-ladenness criteria - *choice space*, *justification*, and *agreement amongst peers* - provide insight into the potential for modeller bias and subjectivity when making assumptions. The 2050 scores, particularly for the criteria *choice space* and *agreement amongst peers*, suggests a much greater opportunity for individual values, biases and subjective judgements to enter the frame, for long term assumptions (Figure 5). This is intuitive; as the knowledge base underpinning assumptions weakens, there is less agreement amongst the expert community about the correct value range or appropriate structural assumptions to employ, and also a wider range of plausible choices that could be assumed in the model.

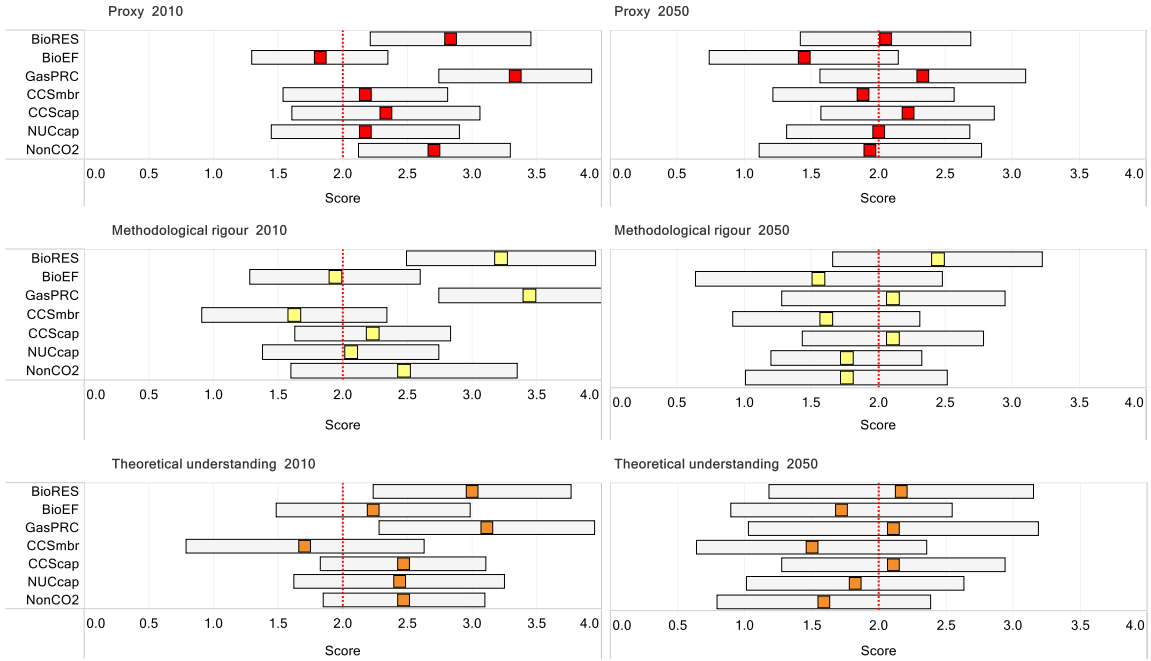


Figure 4. Mean and standard deviation for scores relating to proxy, methodological rigour and understanding criteria. The graphs on the left of the panel are for nearer term assumptions (2010, except 2030 for CCS) and on the right, for 2050. The red line provides a reference point for comparison, indicating the central score value of 2. Descriptions of the mode assumption abbreviations on the vertical axis can be found in Table 1. Scores across the structural assumptions can be found in Annex D (Figure D3).

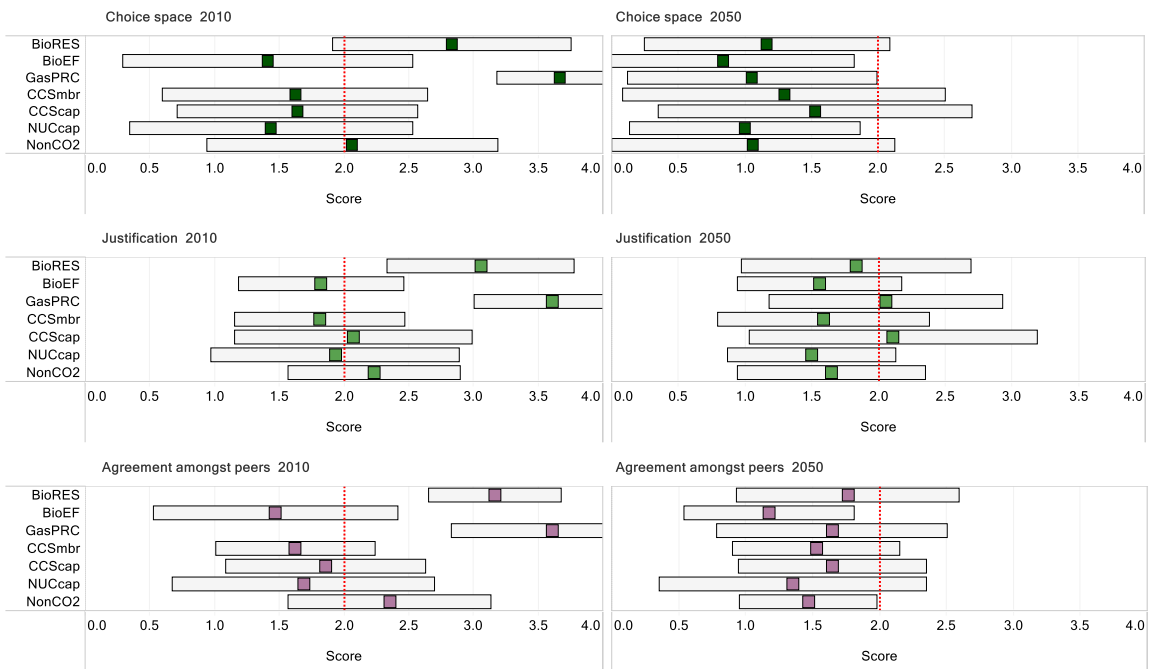


Figure 5. Mean and standard deviation for scores relating to choice space, justification and agreement amongst peers criteria. The graphs on the left of the panel are for nearer term assumptions (2010, except 2030 for CCS) and on the right, for 2050. The red line provides a reference point for comparison, indicating the central score value of 2. Descriptions of the mode assumption abbreviations on the vertical axis can be found in Table 1. Scores across the structural assumptions can be found in Annex D (Figure D3).

3.3 NUSAP diagnostic diagram

As described earlier, the NUSAP approach employs a diagnostic diagram to provide a comprehensive view of model uncertainty by bringing together insights on qualitative dimensions of uncertainty and from quantitative analysis. Figure 6 plots the aggregate pedigree scores, based on the five pedigree criteria as in Figure 2 (horizontal axis), against the sensitivity measure (vertical axis), derived from a Morris Method global sensitivity analysis performed on the ESME model [59], which is further described in Appendix E. The global sensitivity analysis reports the magnitude of the influence on the model results given movement over the full range of possible input values, providing a global measure of influence, or importance. The analysis here broadly replicates that undertaken by Usher [52], one of the two papers used to determine which assumptions to focus on for the NUSAP workshop (as described earlier in Section 2.3). The figure only includes the five data input parameters considered in the sensitivity analysis (see Figure E1, Appendix E).

While the sensitivity analysis shows the strong influence of the CCS build rate assumption (CCSmbr) on the overall costs of decarbonising the UK energy system to hit 2050 climate targets, such an assumption scores weakest across the pedigree criteria. The assumption therefore lies in Q4, or the ‘danger zone’. Such a result matters for decision makers, as the weaker pedigree score explicitly highlights the limits of knowledge underlying this assumption (not necessarily apparent just because an input assumption has a wide uncertainty range), and questions the robustness of any decisions supported by the quantitative analysis. The NUSAP analysis also highlights the large choice space that is possible in the future for this parameter, and limited consensus across experts. Two important conclusions from low scores are that further research and demonstration is needed to enhance the underlying knowledge base for CCS, which is limited, and that decision makers need to be circumspect about developing a strategy underpinned by this yet-to-be commercialised technology. While gas prices and biomass resource level have stronger pedigree scores in 2050, the significant impact of these parameters on the model solution suggests that additional research could help improve the pedigree of these model assumptions further. They perhaps merit stronger attention in future than assumptions about nuclear and CCS capex, which can be seen in Figure 6 to have less of an impact on the results of the sensitivity analysis.

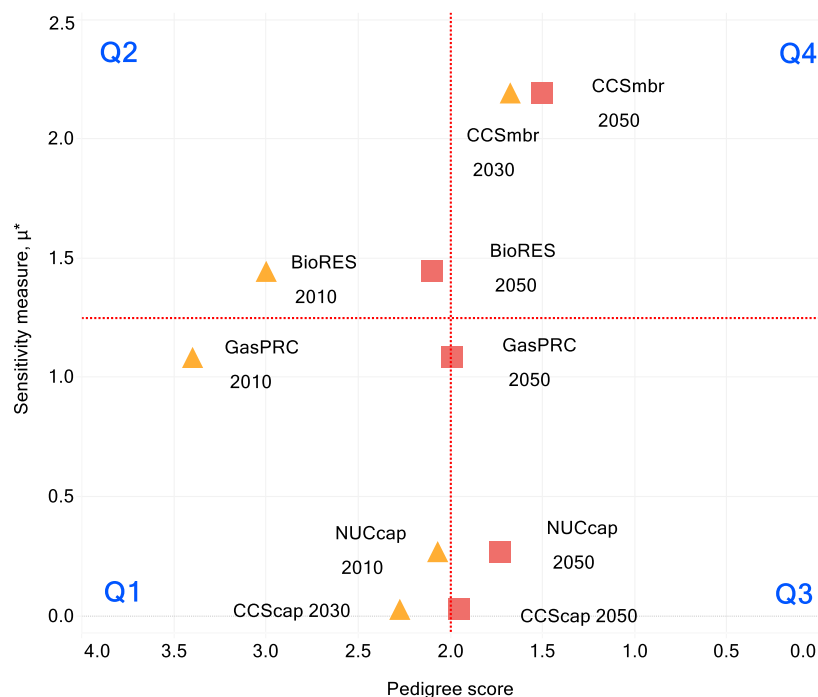


Figure 6. Diagnostic diagram to compare qualitative (pedigree scores) against quantitative uncertainties (sensitivity measure). The sensitivity measure (based on the Morris Method approach) highlights the

influence of the modelled uncertainty on the variance across the model objective function, the total discounted system costs.

The two model structural assumptions (technology learning rates, TchLRN, and perfect foresight optimisation, PerFOR) do not feature in the sensitivity analysis we have undertaken, nor do the assumptions NonCO2 and BioEF. This is because the emissions accounting for non-CO2 GHG emissions (NonCO2) is not an explicit model input assumption while emissions accounting that would capture bioenergy emission factors (BioEF), is not yet featured in the ESME simulation model. On Non-CO2, there is evidence to suggest that this assumption, scoring 1.5 in 2050, could be in or near the danger zone (Q4). Usher [52] considered the impact of changing the UK's emission reduction target, which non-CO₂ GHG emissions would impact, on the results, and found that this is the second most influential model parameter.

4 Discussion and conclusions

The application of the NUSAP approach to energy models such as ESME can help improve the modelling process by making clearer what the knowledge base behind model assumptions actually is, and its strengths and weaknesses. A key benefit of undertaking a NUSAP assessment is that the process specifies for decision-makers *which* model-based parameters are perceived to be particularly weak and exactly *how* they are weak. The diagnostic diagram in particular simplifies large amount of information into an easily digestible graphic for policy stakeholders, highlighting key findings from both the NUSAP workshop and quantitative analysis. Once recognised, decision makers can make better judgments about the robustness of quantitative insights from models, while modellers can seek to improve the transparency and quality of assumptions.

In our assessment presented in this paper, the pedigree scores highlighted which of the most influential model parameters had the weakest underlying knowledge bases. We found that some of the factors with a weak pedigree are those which crucially underpin elements of UK energy and climate policy. These include those on bioenergy (particularly in relation to emission factors), which is viewed as being critical for mitigation in hard to decarbonise sectors like aviation; the treatment of non-CO₂ emissions, which has a direct bearing on the stringency of CO₂ emission targets; and the possible deployment rates for natural gas fired power plant with carbon capture and storage (CCS). For CCS, the deployment assumption is highlighted as being found in the diagnostic diagram's 'danger zone' in Figure . This suggests that decision makers using this type of model should proceed with caution when drawing policy conclusions from model solutions that rely heavily on CCS.

The novel application of a NUSAP assessment that employs differentiated scoring for near and long-term assumptions has highlighted the weaker pedigree for assumptions in 2050, and their greater potential for value-ladenness. For nearer term assumptions, this points to an opportunity for strengthening pedigree through wider consultation with sector experts, development of model structure, and further research across different technologies. It also highlights those assumptions that are well understood, and where there is greater confidence in the knowledge base e.g. for gas prices and current bioenergy resource levels. For assumptions about longer term futures, the analysis highlights a need to recognise that these assumptions are fundamentally different from those relating to the present day, and that strengthening pedigree may not be an option. However, the recognition of a much weaker knowledge base is important in itself, for transparency to users of model outputs, to highlight the need for more research, and also the fundamental need for uncertainty analysis. This includes recognising the socio-political uncertainty that was highlighted by participants when scoring, particularly in relation to long term assumptions on gas prices, bioenergy resources and CCS deployment.

On value-ladenness, the criteria scores highlight that there are frequently many plausible choices that can be made for different assumptions, which often makes the justification of choices challenging, and also that there is sometimes disagreement amongst peers regarding which choices are

appropriate. This confirms the need for more debate on specific model assumptions, for the transparency of choices to be increased, and for broadening the range of opinions to be tested, even those that are perceived as being outside of the ‘mainstream’. Such insights were only gained through having a diverse set of stakeholders, in terms of discipline and organisational background, involved in the workshop scoring exercise.

Of equal interest to the specific findings of the study are the general insights from the process of undertaking the NUSAP workshop. Such a process allowed for extended consultation with a wide range of stakeholders, and revealed the logic and data behind critical model assumptions. While a formal evaluation of the workshop by the participants was not undertaken *ex post*, specific feedback from a range of stakeholders was that the workshop had succeeded in enhancing the understanding of the modelling approach, and the strengths and weaknesses across the underpinning assumptions. This suggests an opportunity for such a process to help engender higher levels of trust in the analytical process, and across the broader expert base. With this type of complex modelling often having been viewed as a black box in the past, opening up the detailed workings to critical scrutiny should contribute to the process of making energy systems analysis more transparent for decision makers [60,61]. This also helps avoid any risks of models gaining authority on or legitimising a specific position based on poor data and the opinions of analysts [62,63].

This constructive dialogue also allowed for participation in the modelling process by outside experts, and enabled the sharing of new perspectives that are often absent from conversations within established modelling teams, or for which the plurality of views are insufficiently recognised. The NUSAP workshop also provided key learning opportunities for those modellers who were present, forcing consideration of the strength of the relationships between elements of the modelled system and the real world system that it is trying to represent, partly as a result of the interactions with experts in specific domains of interest. It also provided the opportunity for modellers to learn about how model input and outputs are perceived and interpreted by others.

The ability of NUSAP to highlight the most critical uncertainties in models can also represent a structured method of selecting the most relevant dimensions for assessment in scenario-based analyses. This brings increased methodological rigour to the often chaotic and ad hoc process that characterise much scenario development activity [64–66] and potentially avoids the all-too-common situation where scenario-and-simulation exercises simply vary 1 or 2 parameters of interest without a thoroughly grounded understanding of whether these are really critical to influencing modelled outcomes.

This first foray into the application of this approach on a national model of this kind has thrown up a number of issues that future research should consider. First, there is an open question as to whether some of the qualitative assessment criteria we employed worked equally well for both model input and structural model assumptions. We observed that many participants who were comfortable using our pedigree scoring system for model inputs were at the same time unsure as to their applicability when it came to discussing model structural features. For example, as noted in the results, participants found it difficult to apply scoring for the *proxy* criterion when considering model perfect foresight, as this is not intended to directly represent real-world decision making. Nevertheless, we would argue that the inclusion of structural parameters in addition to input parameters in our assessment enhanced the overall learning process for our workshop participants by forcing contemplation and discussion of the nature of the model and its relationship to reality in detail. Future work could consider how to integrate a wider set of model assumptions into the pedigree assessment, be they input assumptions, features of the model structure, or broader underlying socio-political assumptions being used in the model analysis. Second, a broader stakeholder base could have been consulted as part of the workshop, including a wider set of disciplines and more domain experts, for example on bioenergy, CCS, and gas markets. However, their absence was not by design but reflected the challenges of workshop organisation.

The NUSAP approach is clearly a time-intensive process that requires the active participation of a group of highly engaged stakeholders, and one that may well push up against the resource and time

constraints of the policy making process. Despite this, we would argue that for major new policy proposals or strategies, using energy models that encompass high complexity and large uncertainties, and which may not be well known, such an exercise is well worth the investment. The analytical process for such proposals will anyway require transparency, broad stakeholder engagement, transdisciplinary framing, scrutiny of assumptions, and recognition of uncertainty, all of which the NUSAP approach covers. Streamlining the NUSAP process to be more light touch would be extremely challenging. While it might be an attractive idea to develop a cut-down NUSAP approach that might fit better with the policy making timetable, we believe that all of the elements described in our example approach are actually needed to strengthen the science-policy interface.

We therefore would recommend that such an approach is integrated into government guidance on policy analysis, recognising the qualitative dimension of uncertainty and an approach to deal with it. For example, in the UK it could be integrated into the formal guidance document for the use of model-based evidence in government policy, the Aqua book [67]. The addition of the NUSAP approach to this guidance would enable practitioners to assess the strength of the knowledge base underpinning their analytical approaches, and reflect any implicit value ladenness in model outputs. If the time-pressure and demands of the policy cycle mean that this cannot be employed on a routine basis, then it should at least be used following the introduction of new analysis techniques to the decision-making process or especially in advance of using models for any major policy decisions. To do otherwise would risk stripping quantitative model outputs of their qualitative context, with the result that decision makers may find themselves flying blind.

In conclusion, the NUSAP approach allows for the recognition of non-quantifiable uncertainties that are often co-produced as part of a highly inter-disciplinary modelling process. Whilst providing a more comprehensive pathway for uncertainty assessment, the approach also enhances the modelling process by facilitating both engagement with outside experts and enabling the necessary scrutiny of model assumptions. Given the importance of uncertainty assessment in modelling for public policy, not only in the UK but internationally, a stronger recognition of non-quantifiable uncertainty, and its inevitable coproduction in the policy assessment process, as well as techniques for addressing it, are all important elements to integrate into formal guidance

5 Acknowledgements

The authors would like to thank the UK Energy Technologies Institute (ETI) who provided the main funding support for this study under the “Energy Pathways Under Deep Uncertainty” project. Funding support for the workshop was also provided by the UK Engineering and Physical Sciences Research Council (EPSRC) under the Whole Systems Energy Modelling Consortium (wholeSEM) [Grant EP/K039326/1].

As well as the funders, the authors would like to extend our thanks to the workshop participants, the advice provided by Jeroen van der Sluijs regarding the research and workshop design, and the participation in the research by Chris Heaton and Scott Milne from the Energy Technologies Institute (ETI) ESME modelling team.

6 Author contributions

S.P. and F.L. designed the research approach, and led the workshop activity, with advice from A.P. These three authors, with W.M. and O.B., ran the workshop, acting as facilitators. S.P. led the analysis of the workshop results, with major contribution from F. L. S.P. and J.P. undertook the global sensitivity analysis, with advice from W.U., who designed the implementation of the approach. S.P. and F.L. led the writing of this paper, with contributions and critical review by all other authors.

7 References

- [1] F.G.N. Li, S. Pye, Uncertainty, politics, and technology: Expert perceptions on energy transitions in the United Kingdom, *Energy Res. Soc. Sci.* 37 (2018) 122–132. doi:10.1016/j.erss.2017.10.003.
- [2] HM Government, *Climate Change Act 2008*, HMSO, London, UK, 2008. http://www.legislation.gov.uk/ukpga/2008/27/pdfs/ukpga_20080027_en.pdf.
- [3] P.G. Taylor, P. Upham, W. McDowall, D. Christopherson, Energy model, boundary object and societal lens: 35 years of the MARKAL model in the UK, *Energy Res. Soc. Sci.* 4 (2014) 32–41.
- [4] N. Strachan, B. Fais, H. Daly, Reinventing the energy modelling–policy interface, *Nat. Energy.* 1 (2016) 16012.
- [5] BEIS, *2015 UK Greenhouse Gas Emissions, Final Figures, 2017*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/604350/2015_Final_Emissions_statistics.pdf.
- [6] G.C. Unruh, Escaping carbon lock-in, *Energy Policy.* 30 (2002) 317–325. doi:10.1016/S0301-4215(01)00098-2.
- [7] S.O. Funtowicz, J.R. Ravetz, *Uncertainty and quality in science for policy, 1990*. doi:10.1016/0921-8009(92)90014-J.
- [8] S.O. Funtowicz, J.R. Ravetz, Science for the post-normal age, *Futures.* 25 (1993) 739–755. doi:10.1016/0016-3287(93)90022-L.
- [9] T.S. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, 1962.
- [10] J.P. Van Der Sluijs, M. Craye, S. Funtowicz, P. Kloprogge, J. Ravetz, J. Risbey, Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System, *Risk Anal.* 25 (2005) 481–492. doi:10.1111/j.1539-6924.2005.00604.x.
- [11] B. Wynne, Uncertainty and environmental learning. Reconceiving science and policy in the preventive paradigm, *Glob. Environ. Chang.* 2 (1992) 111–127. doi:10.1016/0959-3780(92)90017-2.
- [12] W.E. Walker, P. Harremoës, J. Rotmans, J.P. van der Sluijs, M.B.A. van Asselt, P. Janssen, M.P. Kreyer von Krauss, *Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support*, *Integr. Assess.* 4 (2003) 5–17. doi:10.1076/iaij.4.1.5.16466.
- [13] S. Pye, C. Bataille, Improving deep decarbonization modelling capacity for developed and developing country contexts, *Clim. Policy.* 16 (2016) S27–S46. doi:10.1080/14693062.2016.1173004.
- [14] J. Markard, R. Raven, B. Truffer, Sustainability transitions: An emerging field of research and its prospects, *Res. Policy.* 41 (2012) 955–967. doi:10.1016/j.respol.2012.02.013.
- [15] United Nations, *Adoption of the Paris Agreement*, Conf. Parties Its Twenty-First Sess. 21932 (2015) 32. doi:FCCC/CP/2015/L.9.

- [16] C. Bataille, H. Waisman, M. Colombier, L. Segafredo, J. Williams, F. Jotzo, The need for national deep decarbonization pathways for effective climate policy, *Clim. Policy*. 16 (2016) 57–526. doi:10.1080/14693062.2016.1173005.
- [17] G. Holtz, F. Alkemade, F. de Haan, J. Köhler, E. Trutnevyte, T. Luthé, J. Halbe, G. Papachristos, E. Chappin, J. Kwakkel, S. Ruutu, Prospects of modelling societal transitions: Position paper of an emerging community, *Environ. Innov. Soc. Transitions*. 17 (2015) 41–58. doi:10.1016/j.eist.2015.05.006.
- [18] W. Usher, N. Strachan, Critical mid-term uncertainties in long-term decarbonisation pathways, *Energy Policy*. 41 (2012) 433–444. doi:10.1016/j.enpol.2011.11.004.
- [19] S. Pye, W. Usher, N. Strachan, The uncertain but critical role of demand reduction in meeting long-term energy decarbonisation targets, *Energy Policy*. 73 (2014) 575–586. doi:10.1016/j.enpol.2014.05.025.
- [20] S. Pye, N. Sabio, N. Strachan, An integrated systematic analysis of uncertainties in UK energy transition pathways, *Energy Policy*. 87 (2015) 673–684. doi:10.1016/j.enpol.2014.12.031.
- [21] W. McDowall, E. Trutnevyte, J. Tomei, I. Keppo, UKERC Energy Systems Theme Reflecting on Scenarios, UK Energy Res. Centre-UKERC. (2014).
- [22] N. Strachan, R. Kannan, Hybrid modelling of long-term carbon reduction scenarios for the UK, *Energy Econ*. 30 (2008) 2947–2963. doi:10.1016/j.eneco.2008.04.009.
- [23] G. Anandarajah, N. Strachan, P. Ekins, R. Kannan, N. Hughes, Pathways to a Low Carbon Economy: Energy Systems Modelling, UK Energy Research Centre (UKERC), London, UK, 2009. <http://www.ukerc.ac.uk/publications/pathways-to-a-low-carbon-economy-energy-systems-modelling-working-paper.html>.
- [24] N. Strachan, S. Pye, R. Kannan, The iterative contribution and relevance of modelling to UK energy policy, *Energy Policy*. 37 (2009) 850–860. <http://discovery.ucl.ac.uk/170049/>.
- [25] P. Ekins, I. Keppo, J. Skea, N. Strachan, W. Usher, G. Anandarajah, The UK energy system in 2050: comparing low-carbon, resilient scenarios, UK Energy Research Centre, 2013.
- [26] J. Alcamo, Chapter Six The SAS Approach: Combining Qualitative and Quantitative Knowledge in Environmental Scenarios, in: J. Alcamo (Ed.), *Environ. Futur. Pract. Environ. Scenar. Anal.*, Elsevier, 2008: pp. 123–150. doi:10.1016/S1574-101X(08)00406-7.
- [27] P.P. Craig, A. Gadgil, J.G. Koomey, What Can History Teach Us? A Retrospective Examination of Long-Term Energy Forecasts for the United States, *Annu. Rev. Energy Environ*. 27 (2002) 83–118. doi:10.1146/annurev.energy.27.122001.083425.
- [28] E. Trutnevyte, W. McDowall, J. Tomei, I. Keppo, Energy scenario choices: Insights from a retrospective review of UK energy futures, *Renew. Sustain. Energy Rev*. 55 (2016) 326–337. doi:10.1016/j.rser.2015.10.067.
- [29] E. Trutnevyte, Does cost optimization approximate the real-world energy transition?, *Energy*. 106 (2016) 182–193. doi:10.1016/j.energy.2016.03.038.

- [30] F.G.N. Li, E. Trutnevyte, Investment appraisal of cost-optimal and near-optimal pathways for the UK electricity sector transition to 2050, *Appl. Energy*. 189 (2017) 89–109.
- [31] W. Usher, N. Strachan, An expert elicitation of climate, energy and economic uncertainties, *Energy Policy*. 61 (2013) 811–821. doi:10.1016/j.enpol.2013.06.110.
- [32] J. Watson, R. Gross, I. Ketsopoulou, M. Winskel, The impact of uncertainties on the UK's medium-term climate change targets, *Energy Policy*. 87 (2015) 685–695. doi:10.1016/j.enpol.2015.02.030.
- [33] A.C. Petersen, A. Cath, M. Hage, E. Kunseler, J.P. van der Sluijs, Post-Normal Science in Practice at the Netherlands Environmental Assessment Agency, *Sci. Technol. Human Values*. 36 (2011) 362–388. doi:10.1177/0162243910385797.
- [34] C. Heaton, Modelling Low-Carbon Energy System Designs with the ETI ESME Model, 2014. http://www.eti.co.uk/wp-content/uploads/2014/04/ESME_Modelling_Paper.pdf.
- [35] S. Pye, H. Daly, Modelling sustainable urban travel in a whole systems energy model, *Appl. Energy*. 159 (2015) 97–107. doi:10.1016/j.apenergy.2015.08.127.
- [36] F.G.N. Li, S. Pye, N. Strachan, Regional winners and losers in future UK energy system transitions, *Energy Strateg. Rev.* 13–14 (2016) 11–31. doi:10.1016/j.esr.2016.08.002.
- [37] ETI, Options, Choices, Actions: UK scenarios for a low carbon energy system transition, 2015. <http://www.eti.co.uk/insights/options-choices-actions-uk-scenarios-for-a-low-carbon-energy-system/>.
- [38] CCC, Energy system modelling using the Energy Technologies Institute ESME model for the Renewable Energy Review, The Committee on Climate Change (CCC), London UK, 2011.
- [39] J.P. Van der Sluijs, J. Potting, J. Risbey, D. van Vuuren, B. de Vries, A. Beusen, P. Heuberger, S. Corral Quintana, S. Funtowicz, P. Kloprogge, D. Nuijten, A. Petersen, J. Ravetz, Uncertainty assessment of the IMAGE/TIMER B1 CO₂ emissions scenario, using the NUSAP method, Rep. Comm. by Netherlands Natl. Res. Progr. Glob. Air Pollut. Clim. Chang. RIVM, Bilthoven, Netherlands. (2002) 225.
- [40] P. Kloprogge, J.P. Van der Sluijs, A.C. Petersen, A method for the analysis of assumptions in model-based environmental assessments, *Environ. Model. Softw.* 26 (2011) 289–301. doi:10.1016/j.envsoft.2009.06.009.
- [41] N.E. Vaughan, C. Gough, Expert assessment concludes negative emissions scenarios may not deliver, *Environ. Res. Lett.* 11 (2016). doi:10.1088/1748-9326/11/9/095003.
- [42] B.P. Weidema, M.S. Wesnaes, Data quality management for life cycle inventories-an example of using data quality indicators, *J. Clean. Prod.* 4 (1996) 167–174. doi:10.1016/S0959-6526(96)00043-1.
- [43] A. Ciroth, S. Muller, B. Weidema, P. Lesage, Empirically based uncertainty factors for the pedigree matrix in ecoinvent, *Int. J. Life Cycle Assess.* 21 (2016) 1338–1348. doi:10.1007/s11367-013-0670-5.
- [44] J.P. van der Sluijs, The NUSAP Approach to Uncertainty Appraisal and Communication, in: *Routledge Handb. Ecol. Econ.*, Routledge, 2017.

- doi:10.4324/9781315679747.ch29.
- [45] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, *Global Sensitivity Analysis. The Primer*, John Wiley & Sons, Ltd, Chichester, UK, 2007. doi:10.1002/9780470725184.
- [46] A.C. Petersen, P.H.M. Janssen, J.P. van der Sluijs, J.S. Risbey, J.R. Ravetz, J.A. Wardekker, H. Martinson Hughes, *Guidance for uncertainty assessment and communication,* (2013) 34.
<http://pblweb.prolocation.net/sites/default/files/cms/publicaties/550032001.pdf>.
- [47] J.C. Refsgaard, J.P. van der Sluijs, A.L. Højberg, P.A. Vanrolleghem, *Uncertainty in the environmental modelling process – A framework and guidance*, *Environ. Model. Softw.* 22 (2007) 1543–1556. doi:10.1016/j.envsoft.2007.02.004.
- [48] J.A. Wardekker, J.P. van der Sluijs, P.H.M. Janssen, P. Klopogge, A.C. Petersen, *Uncertainty communication in environmental assessments: views from the Dutch science-policy interface*, *Environ. Sci. Policy.* 11 (2008) 627–641.
doi:10.1016/j.envsci.2008.05.005.
- [49] C. McGlade, S. Pye, P. Ekins, M. Bradshaw, J. Watson, *The future role of natural gas in the UK: A bridge to nowhere?*, *Energy Policy.* 113 (2018).
doi:10.1016/j.enpol.2017.11.022.
- [50] CCC, *The Fourth Carbon Budget: Reducing Emissions through the 2020s*, The Committee on Climate Change (CCC), London, UK, 2010.
- [51] ETI, *ESME Data References Book*, 2016. <http://www.eti.co.uk/wp-content/uploads/2014/11/ESME-v4.1-Data-References-Book.pdf>.
- [52] W. Usher, *The value of global sensitivity analysis for energy system modelling*, in: 2015.
https://www.researchgate.net/publication/322721974_The_Value_of_Global_Sensitivity_Analysis_for_Energy_System_Modelling.
- [53] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, Wiley, 2004.
- [54] M.D. Morris, *Factorial sampling plans for preliminary computational experiments*, *Technometrics.* 33 (1991) 161–174. doi:10.1080/00401706.1991.10484804.
- [55] J.C. Refsgaard, J.P. Van der Sluijs, J. Brown, P. Van der Keur, *A framework for dealing with uncertainty due to model structure error*, *Adv. Water Resour.* 29 (2006) 1586–1597.
- [56] E. Cox, *Opening the black box of energy security: A study of conceptions of electricity security in the United Kingdom*, *Energy Res. Soc. Sci.* 21 (2016) 1–11.
doi:10.1016/j.erss.2016.06.020.
- [57] R. Gillard, *Unravelling the United Kingdom’s climate policy consensus: The power of ideas, discourse and institutions*, *Glob. Environ. Chang.* 40 (2016) 26–36.
doi:10.1016/j.gloenvcha.2016.06.012.
- [58] J.P. van der Sluijs, J.A. Wardekker, *Critical appraisal of assumptions in chains of model calculations used to project local climate impacts for adaptation decision support—the*

- case of Baakse Beek, *Environ. Res. Lett.* 10 (2015) 045005. doi:10.1088/1748-9326/10/4/045005.
- [59] J. Herman, W. Usher, SALib: An open-source Python library for Sensitivity Analysis, *J. Open Source Softw.* 2 (2017). doi:10.21105/joss.00097.
- [60] K.K. Cao, F. Cebulla, J.J. Gómez Vilchez, B. Mousavi, S. Prehofer, Raising awareness in model-based energy scenario studies—a transparency checklist, *Energy. Sustain. Soc.* 6 (2016). doi:10.1186/s13705-016-0090-z.
- [61] S. Pfenninger, Energy scientists must show their workings, *Nature.* 542 (2017) 393. doi:10.1038/542393a.
- [62] B. Wynne, The institutional context of science, models, and policy: The IIASA energy study, *Policy Sci.* 17 (1984) 277–320. doi:10.1007/BF00138709.
- [63] B. Keepin, A technical appraisal of the IIASA energy scenarios, *Policy Sci.* 17 (1984) 199–276. doi:10.1007/BF00138708.
- [64] A. Martelli, Scenario building and scenario planning: state of the art and prospects of evolution, *Futur. Res. Q.* 17 (2001) 57–70. doi:10.1017/CBO9781107415324.004.
- [65] R. Bradfield, G. Wright, G. Burt, G. Cairns, K. Van Der Heijden, The origins and evolution of scenario techniques in long range business planning, *Futures.* 37 (2005) 795–812. doi:10.1016/j.futures.2005.01.003.
- [66] M.J. Spaniol, N.J. Rowland, The scenario planning paradox, *Futures.* (2017). doi:10.1016/j.futures.2017.09.006.
- [67] HM Treasury, *The Aqua Book: guidance on producing quality analysis for government*, London, 2015. <https://www.gov.uk/government/publications/the-aqua-book-guidance-on-producing-quality-analysis-for-government>.