

---

# Prospects for energy economy modelling with big data: hype, eliminating blind spots, or revolutionising the state of the art?

Accepted Manuscript: 10<sup>th</sup> February 2019

Published in: April 2019, *Applied Energy*, Volume 239, Pages 991-1002, [doi: 10.1016/j.apenergy.2019.02.002](https://doi.org/10.1016/j.apenergy.2019.02.002)

Francis G. N. Li<sup>a</sup>, [francis.li@ucl.ac.uk](mailto:francis.li@ucl.ac.uk)

Chris Bataille<sup>b,c</sup>, [chris.bataille@iddri.org](mailto:chris.bataille@iddri.org)

Steve Pye<sup>a,d</sup>, [s.pye@ucl.ac.uk](mailto:s.pye@ucl.ac.uk)

Aidan O'Sullivan<sup>a</sup>, [aidan.osullivan@ucl.ac.uk](mailto:aidan.osullivan@ucl.ac.uk)

<sup>a</sup> UCL Energy Institute, Central House, 14 Upper Woburn Place, London, WC1H 0NN, United Kingdom

<sup>b</sup> Institut du Developpement Durable et des Relations Internationales (IDDRI), 41 rue du Four, 75006 Paris, France

<sup>c</sup> Energy and Materials Research Group (EMRG), School of Resource and Environmental Management, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

<sup>d</sup> MaREI Centre, Environmental Research Institute, University College Cork, Cork, Ireland

© 2019. This manuscript version is made available under the **CC-BY-NC-ND 4.0** license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

---

## Abstract

Energy economy models are central to decision making on energy and climate issues in the 21st century, such as informing the design of deep decarbonisation strategies under the Paris Agreement. Designing policies that are aimed at achieving such radical transitions in the energy system will require ever more in-depth modelling of end-use demand, efficiency and fuel switching, as well as an increasing need for regional, sectoral, and agent disaggregation to capture technological, jurisdictional and policy detail. Building and using these models entails complex trade-offs between the level of detail, the size of the system boundary, and the available computing resources. The availability of data to characterise key energy system sectors and interactions is also a key driver of model structure and parameterisation, and there are many blind spots and design compromises that are caused by *data scarcity*. We may soon, however, live in a world of *data abundance*, potentially enabling previously impossible levels of *resolution* and *coverage* in energy economy models. But while big data concepts and platforms have already begun to be used in a number of selected energy research applications, their potential to improve or even completely revolutionise energy economy modelling has been almost completely overlooked in

the existing literature. In this paper, we explore the challenges and possibilities of this emerging frontier. We identify critical gaps and opportunities for the field, as well as developing foundational concepts for guiding the future application of big data to energy economy modelling, with reference to the existing literature on decision making under uncertainty, scenario analysis and the philosophy of science.

## Keywords

Energy modelling; climate policy; energy policy; decarbonisation; energy data; big data

## Highlights

- The application of big data to energy economy modelling is a critical research gap
- Big data has the potential to reduce multiple blind spots in energy economy models
- *Bottom-up hybrid* models are a logical starting point for big data integration
- Moving from *data scarcity* to *data abundance* could completely transform the field
- We explore the research possibilities for revolutionising the state-of-the-art

## 1.0 Introduction

The use of quantitative models to assist decision making in energy policy started in the 1970s [1]. Prior to this, energy use had been observed to increase more or less in lock-step with economic growth, so the links between energy prices and demand were largely left unquestioned. This thinking was overturned by the response of industrialised countries to the 1973 and 1979 energy crises, both of which showed decision makers that there was actually significant capacity within their economies for energy efficiency and switching to alternative fuels. Energy economy models, which were enabled by the rapid rise in affordable computing power around the same time period, were initially developed to explore these options [2]. Since the 1970s the scope of application for such models has grown far beyond their original roots, and they now underpin much of the evidence base for energy and climate mitigation policies at the global level [3], and also long term greenhouse gas (GHG) reduction strategies at the national level [4]. Energy economy models are critical to strategic planning by governments, who are responsible for delivering energy security, economic development and environmental protection within an uncertain and rapidly changing landscape of international relations and natural environmental limits. A particularly challenging example, where models are critical, is decarbonisation planning under the 2015 Paris Agreement [5]. This commits nearly all governments to transition away from the combustion of fossil fuels to atmosphere by the end of the century, and likely by 2060-2080 to stay “well below [a] 2°C” increase from preindustrial temperatures [6].

Since their inception, energy economy models have been designed and developed to operate under conditions of *data scarcity*. Not only are such models used to inform decisions that are taken under future conditions characterised by deep uncertainty<sup>1</sup> [7], and where there is insufficient information to reliably apply formal statistical methods, but the underlying baseline data used to characterise the current energy system are often sparse and contain critical gaps. *Data scarcity*, combined with historical computational limitations, has enforced a reductionist representation of many real world phenomena in models, with available data and tools often driving model structure and dynamics. For example, a lack of information on the spatiotemporal use of energy in the real world and discrete options to mitigate GHGs often leads to the use of representative load curves and idealized mitigation cost curves<sup>2</sup> in engineering models. An example from economic modelling is the widespread use of optimisation and computable general equilibrium (CGE) type models, despite our awareness that real world firms and households are neither perfectly cost minimizing nor narrowly welfare maximizing, and that inputs cannot generally be smoothly substituted in response to price changes, as is common in these models. *Data scarcity* has dictated extensive idealization, abstraction and isolation of complex processes [8], and has

---

<sup>1</sup> Paraphrasing Lempert et al. [150], we define deep uncertainty as the condition in which analysts do not know or the parties to a decision cannot agree upon: (1) the appropriate models to describe interactions among a system’s variables; (2) the probability distributions to represent uncertainty about key parameters in the models, and/or; (3) how to value the desirability of alternative outcomes

<sup>2</sup> The industrial end-use portion of the primary model used by the US government, NEMS, is still constructed this way. See the 2018 NEMS industry sector documentation: [https://www.eia.gov/outlooks/aeo/nems/documentation/industrial/pdf/mo64\(2018\).pdf](https://www.eia.gov/outlooks/aeo/nems/documentation/industrial/pdf/mo64(2018).pdf)

also split the energy modelling world into two fundamentally different paradigms. One is based on the use of *top-down* economic statistics to populate CGE or macroeconomic models, and the other features models built from the *bottom-up*, using information on the quantities, characteristics and usage patterns of buildings, vehicles and machines that transform and consume energy [9]. *Data scarcity* continues to present fundamental challenges to both strategies for applying models to policy problems.

Current innovations in data collection, storage, retrieval and processing have already begun to transform science in a number of areas, notably high-energy physics and astronomy [10]. In the near future, these innovations may offer the potential to change the operating environment for energy models from one that is characterised by *data scarcity* to one that is instead characterised by *data abundance*. However, an extensive review of the literature (introduced in Section 2.0) finds no work addressing the application of so-called big data concepts and tools to energy economy modelling. Accordingly, we find that the energy modelling field has not yet begun to grapple with what future advances in data science might mean for established practices, tools and techniques. In this paper, we explore the potential for big data to contribute to the enhancement of existing energy economy modelling for policy analysis, both in the near future and over the longer-term.

## 2.0 Big data and energy research: the status quo

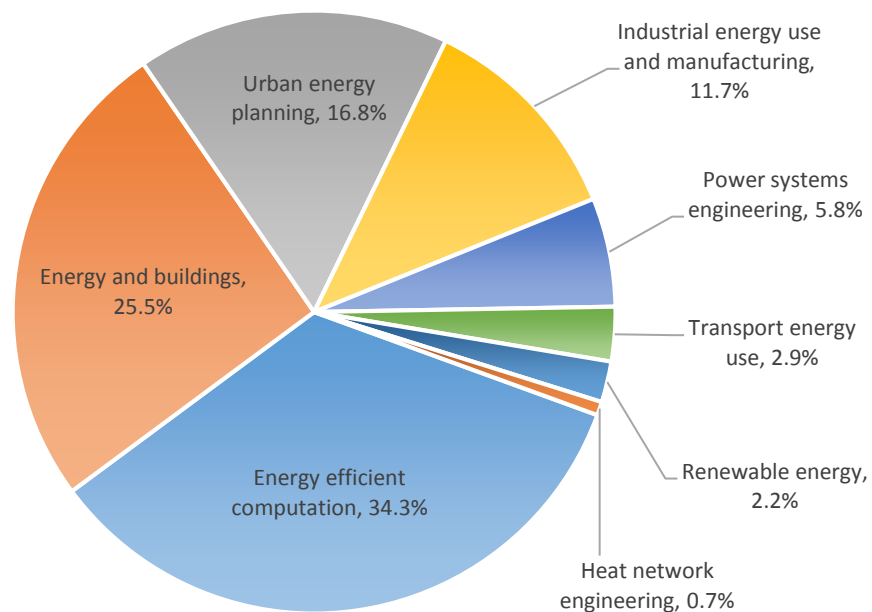
There is broad agreement that information is being created and captured digitally at unprecedented speed and scale [11]. There is however, less agreement on how to characterise this trend, and on what the potential implications are for society, industry, governments, and scientific research. While *big data* is a rapidly emerging field, there are several overlapping definitions of what exactly constitutes *big data* in general use [12]. We will return to our own taxonomy for data in energy economy models later in the paper (Section 3.1), but the contested nature of the concept (i.e. *what exactly is "big" data, and at what size does data become "big?"*) merits highlighting early in the discussion.

As well as being contested in terms of precise definitions, an additional characteristic of the current research landscape for big data is that the scientific community as a whole is arguably lagging behind commercially motivated efforts, with a significant proportion of the innovation, conceptual development, and general discourse on the subject being driven by industry. Much of the existing interest in big data is therefore focused on commercial applications in specific sectors, particularly healthcare [13], through an enhanced understanding of consumer trends [14], visualisation of these trends [15], and analysing the resulting implications for marketing and ecommerce [16].

For this study we carried out a systematic keyword search across the web's major scientific databases, including Scopus, Web of Science, and Google Scholar [17]. For Scopus and Web of Science, which index mainly abstracts and titles, we used the Boolean search term "*big data*" + "*energy model\**", while for Google Scholar, which searches full text entries for published works, this was refined to be more specific; we used "*big data*" + "*energy modelling*". This exercise was initially carried out in March 2018 and repeated again in December 2018. The literature review followed a multi-stage process whereby the initial

keyword searches were subsequently followed by a human review of all papers to evaluate and classify relevant content, as well as filtering out duplicate entries that appeared in more than one database. In total, 283 unique works were analysed. Of these, 146 were found to have the key search terms located only in their reference lists or to mention them only briefly, without detailed analysis or discussion. A quantitative summary of the remaining 137 papers, coded by thematic area, is presented in Figure 1.

Figure 1 – Thematic distribution of reviewed literature, n=137



The existing scientific literature links big data to energy and technology policy in a few specific domains. The largest group of papers explored energy efficient computing, such as techniques for maximising performance-per-watt in while processing large datasets [18] or the potential to reduce energy consumption from datacentres [19]. The second largest group focused on the possible contribution of big data to achieving an enhanced understanding of the energy performance of buildings (e.g. Zhou and Yang [20], Linder et al. [21], Edwards et al. [22]) and the new modelling and decision making possibilities this may enable [23]. The third largest group of publications discussed big data in the context of “smart” cities and urban energy planning, where the collection of large datasets from smartphones and autonomous sensors is an oft-cited example of how a step-change in data availability might transform policymaking and city governance (e.g. Batty [24], Rathore et al. [25]). The fourth significant group investigated the potential of big data for sustainable manufacturing and reducing industrial energy use [26], including specific studies on individual processes like steelmaking [27] and machining [28]. Taken together these four large thematic categories comprised 88% of the review sample.

The remaining 12% of the sample reveals several interesting use cases for big data and energy research. One of these is the electricity system, where scientists are investigating

the possible design configurations for future “smart grids” and exploring the benefits for system management, stability, and operational flexibility that might be achieved (e.g. Alahakoon and Yu [29], Akhavan-Hejazi and Mohsenian-Rad [30]); this could be important in future grids that combine intermittent renewable generation (both distributed and centralised) with other generation types. Another is a nexus of big data research centred on the analysis of complexity in transport patterns as a means of anticipating how they might shift or be managed in future (e.g. [31]). The two final areas of interest that appeared in the review were the application of big data techniques to explore renewable energy potential [32] and the design and operation of heat networks [33].

There is clear evidence therefore, that big data platforms and concepts have already begun to be applied in various sub-domains of the energy system. However, our 200+ paper review finds that the application of modern data science specifically to energy economy modelling and the implications for strategic decision making on energy and climate mitigation policy remain entirely unexplored. A recent meta-review of big data research covering the period 1996-2015 also analysed over 200 relevant papers [34], but similarly, found no examples which discussed the future prospects and challenges posed by the emergence of big data for energy economy modelling. We find that highly cited, forward looking perspectives on the potential application of big data to problem solving in industry and in scientific research (e.g. [35]) and many reference texts (e.g. [36]) also continue to overlook this critical area. This means that there is no precedent for applying big data research concepts to the energy economy modelling domain, no guidelines for best practice, and no research outlook to identify priority areas for future investigation. Clearly, it is prudent for energy economy modellers to think ahead, anticipate, and plan for a near-future environment where data may be much more available than it has been in the past. It is the hope of the authors that our efforts in this paper go some way towards bridging this gap.

### **3.0 Foundational concepts**

#### **3.1 Data for building energy economy models**

The term *big data* is found to have multiple definitions when discussed by various industrial and research groups. A number of attempts to define big data are driven by commercial speculation rather than scientific enquiry, and there is a lack of consensus on a universal definition [37]. However various taxonomies in the literature attempt to incorporate *volume* (quantity of information), *variety* (heterogeneity or complexity of information), *velocity* (speed of information production and dissemination), *value* (novelty of information), and *veracity* (quality or pedigree of information) as key characteristics of big data [38]. These are styled in the literature using “\*V” or “the \*V’s”, where “\*” refers to the number of dimensions i.e. “3V” or “the 3V’s of big data”. This has become enough of an established trope to the extent that there are now even papers that discuss up to 10 different “V” dimensions [39].

This approach is potentially problematic because different taxonomies adopt different interpretations (e.g. what *veracity* or *value* really means), and because some descriptors, such as what size constitutes “big”, are relative and may evolve over time [40]. For our

purposes here, we have therefore steered away from creating a taxonomy that attempts to have more (or better) “V’s” than other research efforts. Instead, we are focused on exploring how *data scarcity* or *data abundance* might affect how models are designed and used in energy economics, especially for energy and climate policy analysis. This requires first developing a framework to highlight exactly how and for what purposes data are commonly used in energy economy models today, before progressing to discussions about how radical shifts in data availability might change the established paradigm. For the purposes of informing our discussion here, we make the distinction between four major categories of information that are used in building energy economy models, summarised in Table 1:

Table 1 – Information for building energy economy models

#	Category	Description	Human Input	Example for typical <i>top-down</i> energy models	Examples for typical <i>bottom-up</i> energy models
0	Raw Data	Unprocessed, information from automated sensors or accounting systems	Minimal other than initial configuration	Electricity market transaction records Import/export records showing value of commodities by individual shipments	Electricity meter readings Electronic traffic monitoring records Weather monitoring station readings
1	Primary Data	Empirically grounded data on observed real world quantities, inventories, and physical assets, held in structured databases	Error checking, “cleaning” of raw data, input to structured databases, summary statistics i.e. “ <i>models of data</i> ” [41]	Annual, seasonal or other time-indexed values of fuel imports and exports (\$) Econometrically derived estimates of behavioural variables over time (e.g. price elasticities)	Annual summary of physical assets (e.g. millions of vehicles) Up front capital, labour, material, and energy input costs. Wind and solar energy availability over different time periods
2	Secondary Data	Data derived from multiple Primary Data sources	Often the inputs into energy economy models are actually simulated outputs from other models. Energy modellers must often combine empirical data with conditional assumptions or theory in order to obtain useful parameters.	Diffusion or adoption curves for different technologies based on multi-criteria decision models derived from surveys of consumer preferences and assumptions about the psychology of decision making (bounded rationality, utility maximisation etc.)	Transport activity patterns (e.g. passenger kilometres by population segment) based on data on fuel use and vehicle efficiency data combined with assumptions about average trip distances.
3	Projections	Model inputs which represent analyst assumptions about possible or plausible future system conditions	In the absence of empirical data about the future, model builders often test plausible ranges of future values informed by expert elicitation and judgements [42].	Future economic growth (GDP) over time The rate of autonomous energy efficiency improvement (AEEI) [43] from innovation and capital stock turnover in the energy system	The rate of future growth in households The date of first commercial availability for a future technology

Energy economy models often contain hundreds of different parameters. *Data scarcity* in many areas presents particular challenges for model builders. From a practical standpoint, *data scarcity* means that modellers must often construct models of the energy system to use whatever data are available, and then fill in the blanks as best they can, often relying on expert judgement (Category 3) or evidence from prior studies using other models (Category 2). As with most applied models in science [44], this means that energy economy models typically represent a hybrid between observed data and theories about the functioning of the world. *Data scarcity* also imparts a great degree of uncertainty into model based analysis, because in addition to dealing with simple unknowns and quantifiable risks that arise about the future, modellers must also contend with fundamental uncertainty about how well their models represent the real world in the first instance. A comprehensive review of the literature on characterising these kinds of uncertainties is beyond the scope of this piece, but a range of useful concepts for grappling with these challenges can be found in the work of Knight [45], Wynne [46], and Stirling [47] to name a few.

### 3.2 **Data, models, and reality: what can be achieved with “more” or “better” data?**

Given its potential for making the real world intersection of energy supply and demand and the economy more directly observable and measurable, the availability of big data and conditions of *data abundance* could force energy economy modellers to confront their basic assumptions about what their models represent and how useful they are for policy analysis. To quote a seminal paper by Boyd and Crawford, “*Big Data reframes key questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and the categorization of reality*” [48]. Addressing these issues in a comprehensive fashion requires leveraging the existing literature on decision making under uncertainty, scenario analysis and the philosophy of science.

From a philosophical standpoint, using complex system models for scientific enquiry is often described as an exercise in conducting thought experiments [49] on a reduced representation of reality in order to infer useful knowledge about reality itself (see the work of Frigg [44], Hughes [50], Parker [51] and Mäki [52]). Much in the same way that past explorers made maps, which are themselves abstract approximations of real terrain, to make navigational choices, energy economy models can provide an overview of the strategic decision space upon which energy policy operates [53]. We call this the *models-as-maps* analogy, which has come to feature strongly in the literature on energy decarbonisation pathways [54]. In the same way that the quality of a map affects the ability of an explorer to chart a course through an uncertain landscape, the quality of an energy model influences the decision maker’s understanding about the dynamics and available degrees of freedom that might exist in the future energy economy.

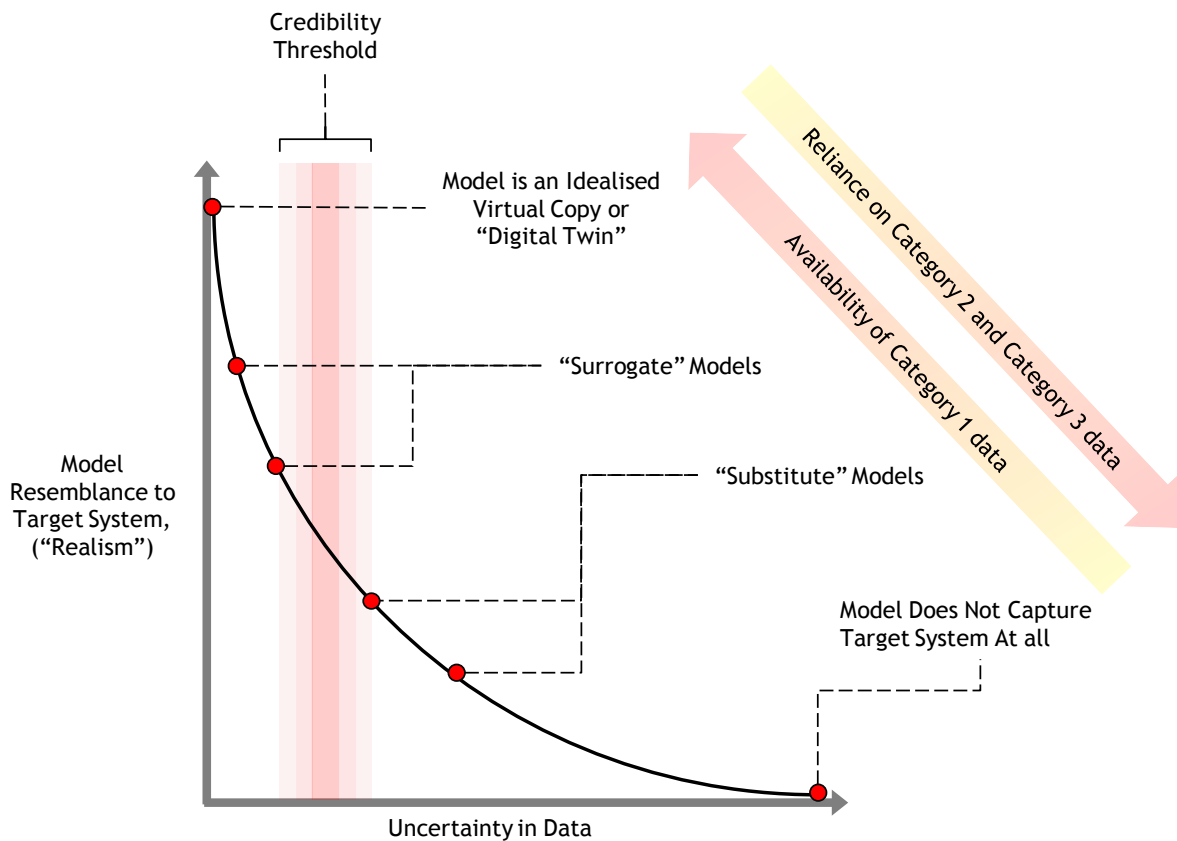
At the deepest level, our understanding of data is comprised by the relationship between our theories of how the world is structured and functions, which determines what we consider worthy of measurement, and our capability to measure data. Clearly, any given model component, such as an oil price, can be characterised with information from the categories in Table 1. But as modellers move along the scale from Category 0 data (raw information) to Category 3 data (projections), an increasing reliance on human value judgements is required. This increases the potential for bias, and consequently, the need to



draw conditional inferences from the model outputs, i.e. more and more model-derived observations can only be said to be true if all other theoretical dependencies input by the model operator (including their opinion and judgement calls) are also held to be true [55]. There are therefore fundamentally different epistemic (or knowledge) claims that can be made on the basis of models built from primary data (Category 1) as compared to models that are built from the analytical outputs of other models (Category 2), or from projections driven strongly by theory or analyst opinion (Category 3). Much of the promise of big data as conceived by its proponents lies in the potential for the increased availability of high resolution, granular information on a huge range of areas. In our taxonomy from Table 1, this could mean a large increase in empirical data (Category 1).

We cannot know in advance exactly what future advances in big data might bring for characterising energy economy models, although we do discuss some of the possibilities in the next section (Section 4.0). We can however, conceive of various hypothetical states in terms of the relationship between data quality, systemic uncertainty and model realism. A stylised depiction of this relationship is featured in Figure 2. On the vertical axis we illustrate the resemblance of the model to the real world, while on the horizontal axis we depict the level of systemic uncertainty in the data used to create the model. We adopt here Walker et al.'s definition of uncertainty as "*any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system*" [56]. We conceptualise the epistemic power of models as being somewhere on a spectrum between not capturing the target system at all, at one extreme, and being a virtual copy of reality at the other, a so-called "digital twin" [57]. For the moment, we will leave aside the issue of whether a *digital twin* of a huge and complex system like the energy economy is feasible or not.

Figure 2 – data, uncertainty, and model realism



As well as the two extreme states of (1) perfect replication and (2) a model of negligible utility, we also show various intermediate states. To describe these, we rely on Mäki's taxonomy of model realism [58], and distinguish between those models that can only serve as *substitutes* for reality, and those which qualify as *surrogate* representations of reality. Mäki characterises *substitute* models as those that are disconnected from the real world but which nevertheless may be used as test beds for developing novel techniques. *Surrogate* models, on the other hand, are close enough to reality to be potentially useful for decision making. In philosophical terms, these models are said to provide epistemic access beyond themselves and can serve as a bridge to understanding the real world system (as discussed for example, by philosophers such as Godfrey-Smith [59] and Frigg [60]). *Surrogate based analysis* is a common concept in advanced engineering design [61], and the term is now also entering the language of complex economic modelling [62].

But what distinguishes a *substitute* model, useful only for testing theories, from a *surrogate* model, useful for politically robust decision making? The pragmatic response found in much of the literature is that models start to be used for decision making when the audience is sufficiently convinced by the resemblance of the model to the real world system for the specific problem being investigated [63]. The resemblance between the model world and the real world is therefore critical to this process [64], with its importance not only observed to be true in economics but also in other applied research fields like biology [65]. In Figure 2 we therefore postulate that there is a *credibility threshold* beyond which models resemble reality well enough to be used by decision makers as *surrogate* systems, useful for productively informing real world decisions.

We argue that the main vector by which big data could transform energy modelling in practice is to enable their construction using more primary data (i.e. Category 1), and reducing their reliance on Category 2 or Category 3 information. By moving the modelled system representation to one which more closely approximates reality, model findings might be viewed as more credible, salient and legitimate [66] for the purposes of decision making. This could take models over the *credibility threshold* across a wider range of sectoral, national, sub-national, and policy analysis areas with fewer caveats or warning labels attached to their outputs.

#### 4.0 Applying big data concepts and platforms to energy economy modelling

##### 4.1 Data gaps and opportunities

Table 2 gives an overview of some of the blind spots in existing energy economy models that are caused by *data scarcity*, together with examples of contemporary practice and the challenges that this raises for policy research. We first return briefly to the *models-as-maps* metaphor originally introduced in Section 3.2. While modern maps are in most cases excellent 2-dimensional reproductions of physical geography, we can highlight that this was certainly not always the case. Historically, some maps were incomplete, others were vague on detail, and some even contained incorrect or misleading information. Maps improved over time, with new information allowing map makers to show details that were previously not there, and also to represent regions that were previously unmapped. We can think of big data then, in terms of its potential for:

- a) Increasing the *resolution* of energy models, resolving energy, material and financial flows at hitherto impossible levels of detail, or;
- b) Widening the *coverage* of such models, expanding system boundaries and enabling the characterisation of energy economy phenomena that were previously either uncaptured or only handled in an extremely abstract fashion

Table 2 – Typical blind spots in existing energy economy models caused by *data scarcity*

Data Types	Status Quo	Examples of Typical Contemporary Practice	Challenges
Spatial	Insights typically focused at national or global scales, with highly aggregated geographical coverage. Some models are multi-region but only representing large regions e.g. USA, China, India, Europe in global models.	The UK's ESME model has 21 onshore and offshore model nodes, each representing areas administered by sub-national governments or former administrative regions [67]  The TIAM-UCL model employs 16 regions [68] and for example aggregates all of Africa together.	Spatial aspects of energy transitions (infrastructure, location of resources, local decision making, socio-cultural differences) are difficult to represent without detailed spatial disaggregation.

Temporal	Typically, models use a number of representative diurnal and/or seasonal time periods within each projected future year to characterise dynamic phenomena like changing demand or supply conditions.	The Irish TIMES model uses 12 time slices to characterise 3 diurnal periods (night, day, peak) and the four seasons [69]. Only a limited number of models have been used with increased temporal characterisation (see [70] or [71]).	The low temporal resolution of models makes it difficult to explore the implications of high variable renewable systems, demand response policies like time of use pricing, and the role for energy storage.
Technological	Typical practice is to aggregate the engineering detail of different types of physical stocks (buildings, power stations, vehicles, industrial plants etc.) into broad categories of representative technologies.	The UKTM model [72] represents offshore wind plants (which differ in age, size, design, operation, manufacturer etc.) using a single reference plant with average performance and cost characteristics although differentiated depending on resource area.	Technology aggregation can lead to simplification of the existing energy asset base, and not adequately differentiate on cost / performance grounds between new technology investments.  A good example is that energy economy models represent poorly new tech such as smart appliance, how they may be networked, and their operation.
Socio-economic and demographic	Many models employ an abstract representation of an “average” person per country, which can hide significant differences in energy use across socio-economic groups e.g. income, age, household size, behavioural characteristics etc.	Most <i>bottom-up</i> energy economy models, unlike a number of <i>top-down</i> CGE models (e.g. [73]), do not distinguish different households by income brackets or other differentiations in population energy use.	Different social characteristics are all linked to varying patterns of energy use (see Lamb et al. [74] and Crespo [75]), and it is challenging to capture this in models without the data to study or characterise these relationships. It is also challenging to explore the distributional impacts of different policies (e.g. [76]).
Actor behaviour and preferences	Decisions made by individuals are typically simplified in models to reflect neoclassical economic assumptions about demand and supply, without integrating more complex insights from fields such as behavioural economics, which are known to be critical [77].	The widely used TIMES model family employs cost minimisation or welfare maximisation as simple decision rules to explain how different options are traded-off against one another [78].  An emerging class of experimental model variants, such as MESSAGE-Transport [79], is seeking to better characterise consumer	Past studies reveal that changes in assumptions about behaviour can have large effects on the outcomes of energy transition pathways [80]. However, models typically have limited data on consumer preferences.

		preferences in specific sub sectors	
Type and condition of existing stock (physical assets)	In many countries there is a fundamental lack of data to characterise the energy system in many sectors [81]. This is typically the case in countries with limited resources to develop and maintain statistical databases.	In most circumstances, a model will be built based on an existing annual energy balance, with characterisation of the technology stock then fitted to the energy balance and furnished via other datasets e.g. car tax records, utility bills, building stock registers etc.	A lack of detailed data can make it challenging to target policies and resources at specific stocks e.g. improving conditions for vulnerable energy users in poorly insulated homes, retrofitting old and polluting industries, or targeting older and less efficient vehicles.

Typically, across all data types, it is the power sector and other utilities that hold the most comprehensive and detailed datasets. This particularly the case in those countries with competitive power pools, because a transparent market structure is needed to limit the potential for market manipulation. The high quality datasets for power and other forms of energy production (with the key exception of self-consumption of product, which is considerable in some sectors) usually makes the supply side of the energy system straightforward to characterise when building models. In contrast however, the resolution of time series information and cross sectional data density for representing energy end-use demand can often be quite poor, even in advanced economies. While many countries have registers or databases on their domestic building stock, it remains difficult to characterise micro-scale patterns in household activity [82] in terms of the technologies used, the time of use, and how these differ across regions and demographic groups. The same is true in the commercial and industrial sectors, where energy use is extremely heterogeneous, data availability is typically much poorer, and where there are few centrally held datasets in most countries.

Big data potentially offers multiple opportunities for increasing the *resolution* and *coverage* of information for characterising energy models. Potential sources include real-time feedback from electric power grids [83], mobility data on travel patterns from public metro transit authorities or private transportation network companies [84], smart utility meters in the residential sector [85], advanced databases on buildings [86], satellite imaging [87], and social data from the world wide web [88]. In private road transport and commercial freight, smartphones or embedded networked sensors could potentially serve as data acquisition tools under the internet of things paradigm, which is a concept that is itself facilitated by big data technologies and concepts [89]. High spatial and temporal resolution data could potentially give decision makers a clear idea of how energy is being used and produced at a very granular level with very fine time slices. This could unlock the door to multiple possibilities for policymaking, possibly even in real time as the incidence is imposed. Table 3 gives an overview of potential future big data sources, their application to energy economy models, and examples of current gaps that can be addressed.

Table 3 – Potential big data sources for addressing gaps in models

Big data sources	Modelling application	Examples of current gaps that can be addressed
Smart meters, intelligent grid management systems, non-intrusive load monitoring [83]	<p>Exploring grid flexibility and the impact of smart systems and demand-side management</p> <p>Improved representation of micro-scale patterns in energy use</p> <p>Targeting efficiency policies</p> <p>Impacts of socio-demographic changes on energy consumption.</p>	<p>Timing, type, and level of energy use [90]</p> <p>Technical – quantified, high resolution, understanding of the contributing factors to overall home energy consumption, the contribution of appliances and potential for adaptation [91]</p> <p>Understanding the variation in energy use across population groups and the impact of shifts in population composition [92], insights into socio-demographic factors on energy use [93], willingness-to-pay for energy [94]</p>
Distributed ledgers, e.g. blockchains [95]	Distributed community energy	Energy trade between “prosumers” [96]
High resolution spatial data from satellite imaging, cubesat (miniaturised satellites) sensor data, LIDAR data	<p>Improved emissions, air quality and pollution modelling, such as fugitive methane and oil and gas extraction auto-consumption (self-use)</p> <p>Quantifying weather dependent renewable energy potential e.g. rooftop solar</p> <p>Improved land use tracking and modelling over time</p> <p>Improved resource potential modelling e.g. bioenergy</p> <p>Improved spatial and geographical representation of population and energy consumption</p>	<p>Identification and quantification of true power plant emissions [87] or geospatially granular fugitive methane monitoring e.g. from: <a href="https://www.ghgsat.com/">https://www.ghgsat.com/</a></p> <p>Identification of electricity generation options and viability [97]</p> <p>Air quality and pollution tracking [98]</p> <p>Data source for developing countries, urban population concentration, access to energy [99]</p> <p>3D city modelling, large scale urban modelling [100], building stock modelling [101], and urban environmental accounting tools (e.g. Google’s Environmental Insights Explorer: <a href="https://insights.sustainability.google/">https://insights.sustainability.google/</a>)</p> <p>Agricultural and forestry stocks [102]</p>
Personal mobility data from: mobile phone location records, metro transit card data, traffic video analysis	Improved modelling of travel patterns, modal choices, and costs	<p>Improved understanding of socio-economic factors influencing transport patterns and energy consumption and their evolution</p> <p>Understanding trip distances by purpose and mode e.g. origin-destination matrices [24]</p> <p>Understanding human mobility [103] and migration patterns [104]</p>

Web site and social media data: e.g. Twitter, Facebook, newspapers, comment pages	Textual analysis and quantification of societal trends and public opinion  Characterisation of policy and decision-making environment	Perceptions and attitudes to energy technologies such as CCS [105]  International sentiment and debate on climate change [88]  Quantifying uncertainty regarding the policy environment [106]
---	---	---

The increasing availability of these sources of data, combined with advances in machine learning and artificial intelligence that enable insights to be extracted from them (e.g. cluster analysis [107]), potentially opens up new horizons in the energy modelling landscape. For example, the vast quantities of smart meter data which are becoming available offer the opportunity to build models based on energy consumption patterns from a population of heterogeneous groups [92] rather than relying on a single “average” usage pattern. This could not only improve demographic, spatial and temporal realism but also the ability of models to quantify policy interventions such as demand side response measures (e.g. optimising vehicle charging [108]). New data sources also potentially enable policy to go deeper than simply the household level through the disaggregation of household energy consumption profiles into usage by individual appliances [109]. This could allow for more granular models which could explore the impact of energy efficiency policies targeted at the appliance level.

Satellite imagery is a globally available resource which is finding increasing uses as a means of addressing a number of questions for which accurate data is often hard to come by. In developing economies it is becoming an invaluable tool for accelerating data gathering, identifying energy resources for development [97] and assessing levels of energy access [99]. The large-scale global trend towards increased urbanisation [110] makes cities an ever more vital application area for energy models and data, and increases the need for an improved understanding of urban-scale issues such as the travel patterns of city inhabitants [24] and quantifying the extent and condition of the existing building stock [111]. Finally, advances in text-based analytics are beginning to enable more abstract information to be drawn from internet sources such as newspaper and social media websites. This can be used to assess and quantify estimates for a range of issues such as public opinion towards technologies [105], the relative levels of acceptance of climate change science in different countries [88] and even the levels of uncertainty surrounding economic policy [106].

#### 4.2 Integrating big data into bottom-up hybrid energy economy models

In this section, we identify a specific family of energy economy models which potentially have the most to gain from future conditions of *data abundance*. Energy economy models are complex mathematical constructs that aim to represent the key relationships that drive energy demand and supply, including dynamics in household end-use demand and demand by firms, energy transformation, and primary supply. This typically includes distinct, linked sub-modules depicting activity in residences, commercial buildings, personal and freight transport, industry, and energy supply and transformation facilities (like power plants or refineries). As mentioned briefly in the introduction, there have historically been two broad

schools of thought used for characterising the energy system since the 1970s, which are typically referred to as being the *top-down* and the *bottom-up* approaches [9]. Both techniques rely on different kinds of data to build a representation of the real world system.

*Top-down* models (e.g. MIT-EPPA [112], MRN-NEEM [113], etc.) are built primarily from national economic statistics and macroeconomic projections, including input-output tables, labour force volume and productivity forecasts, and mandatory reporting of accounts such as fuel sales. *Bottom-up* models on the other hand (e.g. MARKAL [114]), tend to rely on physical observation of engineering systems, such as government records or monitoring by utilities on the location, condition and characteristics of power plants, buildings, and vehicles. Both approaches have their strengths and weaknesses. In general, *top-down* models provide a better understanding of policies that drive changes in economic structure and are sensitive to monetary flows, while *bottom-up* models are much better at depicting the effects of technology orientated regulations or the impacts of mixed regulatory and pricing strategies.

The energy policy landscape in most countries has become more complex and increasingly fraught with uncertainty over time. To chart a course towards 2°C compliant futures, there is a need for modelling systems that can represent the complex interactions between command and control approaches and more market-driven, performance based technology regulations, energy and carbon pricing (and associated revenue recycling). At the same time, there is a need to capture the granular details of differentiated sectoral policies and responses, all while also still being able to represent general economic effects in areas like GDP, employment, and trade. These demands have greatly increased the need for strategic analysis using *hybrid* energy economy models.

*Top-down* macroeconomic or CGE models have had their technical detail progressively increased so that they can be used for energy policy purposes such as investigating issues around security of energy supply, air quality, and GHG emissions, in a process called *top-down hybridisation*. Simultaneously, researchers have grafted technically detailed *bottom-up* models of physical energy infrastructures (e.g. buildings, transport, the power sector etc.) on to an overarching understanding of how the economy functions so that changes in pricing, demand and feedbacks from resource scarcity can be explored. This process is known as *bottom-up hybridisation* [81].

*Top-down hybrids* have proven challenging to construct because their general equilibrium structure forces a simplification of decision making agents. Most models make do with one aggregate household that maximizes welfare, while sector firms maximize profits using a stylized “production function” to represent technology in an environment where returns-to-scale are constant. To add transformative technological detail, new sectors or variants on the existing production functions must be added in a way that reflects a discrete shift from the existing production function [115]. However, once users try to incorporate a large number of cumulative policies, sectors, regions, and transformational technologies into the analysis (e.g. alternative low GHG electricity generation technologies [116]), this approach eventually creates computational challenges that are difficult to surmount. To model significant transformations in the energy economy system, such as that required by 2°C compliant decarbonisation, *top-down* models require the addition of capital vintaging and typically an almost complete duplication of their production function structures (which are



largely calibrated to existing technologies), with alternative low carbon equivalents, or the large scale modification of their existing input substitution structures (with elasticities that change through time, etc.). They also face a fundamental challenge when attempting to account for sector behaviour that does not maximise household welfare though maximizing consumption and leisure, or maximising firm profits [9].

We believe that the most promising pathway for the future development of next generation energy economy modelling is to employ big data with *bottom-up hybrids* [81]. Examples of this emergent energy system model family include BLUE [117], CIMS [118] and IMACLIM [119]. These models combine most of the technical detail of *bottom-up* models with much of the behavioural realism of *top-down* models, e.g. CIMS is designed to handle intricate mixes of real world technology regulations and energy/carbon pricing while clearing energy and end-use markets, while IMACLIM does all of the above while going one step further and also adjusts the long run structure of the economy in response. All *bottom-up hybrids* are intended to represent decision making by actors in response to policy action, directly where it is made in the economy, be it by firms, households, or government; in this sense, they are agent based (though some prefer the term "actor based", see [120]). They can capture the conditions found in sub-optimal second-best policy environments where there are market externalities and political barriers to action [121], and can be used for measuring the difference between "*cost optimal*" (i.e. neoclassical) and "*revealed*" behaviour [122]. The increased *resolution* and *coverage* afforded by big data could in the future enable these tools to provide an extremely detailed representation of the real world system, perhaps even approaching that of a convincing *surrogate* model in many areas (Section 3.2).

Transforming a *bottom-up hybrid* into a convincing *surrogate* model would require very large amounts of revealed empirical data, which would differ highly by sector. As the name implies, *bottom-up hybrids* build up their representation of the energy system from detailed sectoral sub-models, representing decision making at the point of investment and operation. As a result of this granular-scale detail, they rely on a lot of technical data as well as behavioural parameters (e.g. decision specific discount rates [123], end-use demand elasticities [43], non-financial welfare costs for modes [124] or technologies [125]). While open source databases can be built for technical performance and equipment pricing data relatively quickly from manufacturing catalogues, behavioural parameters are notoriously onerous and expensive to gather. This is typically achieved through stated and revealed choice studies [123] in specific areas. For example, looking at household purchasing behaviour for heating systems [126] or transport mode choice [127]. These are complicated, time consuming, and expensive to carry out. The onerous data requirements for *bottom-up hybrids* are one of the key impediments to the construction of these models. Near future advances in big data collection (Section 4.1) might offer simpler, more effective, and more comprehensive approaches to gathering the necessary behavioural data for their construction and calibration, potential spurring their more widespread adoption.

## 5.0 Discussion

### 5.1 Reasons for caution

While the authors acknowledge the potential of big data technologies to challenge some of the long standing limitations of energy economy models, we are not blind to the potential barriers and pitfalls that face practitioners working in this area. First, there are critical questions to be addressed regarding the costs of data collection and the ownership of the resulting information. Data is central to the modern information economy and is highly prized by private companies who trade on their business intelligence or sectoral expertise. The incentives to maintain private for-profit databases are therefore extremely large. Accordingly, access to this data is unlikely to come at low costs to governments or researchers unless the data owners are effectively compensated. Government may be the only actor with the necessary size and agency to make large strides towards more open data in the public interest, but careful thought is required to understand how to implement large-scale data collection for the public good in order to achieve mutual societal and environmental benefits without violating individual privacy rights. The negative implications for society of data-related scandals such as that which discredited Cambridge Analytica in 2018 are all too apparent, and studies show that concerns over unauthorised access to personal data have been growing over time [128].

The feasibility of directly measuring critical parameters from the population without violating their privacy or intellectual property is an important area for future research. A common existing technique is the re-aggregation of data for anonymization with monitored access over encrypted storage [129]. However, if these primary data are collected and stored in their original granular form then the risk of a data breach leading to re-identification of individuals may never be reducible to zero. It is apparent that traditional privacy and access control techniques that form part of established industry practices may be inadequate for big data warehousing [130], so clearly more work is needed in this area. Ultimately, the issue of balancing risks to individual autonomy against the potential for wider societal benefits from data collection is an issue not only for public and private research institutions but also for politicians, lawmakers and civil society to collectively resolve [131].

As noted in Section 4.2, the data used to construct complex next generation energy economy models must come from multiple sources and is likely to be highly heterogeneous. As data size alone is no guarantee of quality or utility [132], most big data must be specially prepared before it can be used [40]. The availability of big data on energy systems phenomena therefore not only creates new opportunities for increasing the *resolution* and *coverage* of energy models (see Section 4.1), but also creates a need for new tools and analytical disciplines to be brought to bear [10]. These include methods for revealing the underlying structures, patterns and relationships in data, such as data mining [133] machine learning [134], and advanced visualisation [135]; as well as combining the resulting multi-source, multi-relational information together into federated or composite databases, in a discipline known as data fusion [136]. Notable approaches for data fusion and integration include data warehousing [137], data virtualisation and “data lakes” [138]. The skills to make sense of massive heterogeneous datasets are currently in short supply relative to the potential demand [139], and may not be widespread in the energy systems modelling community. If the future emphasis in energy economics begins to shift away

from theoretically grounded models that are populated with scarce data, and towards analysis that relies more on the discovery and exploration of useful patterns in abundant data, then a major priority for research will be skills development and capacity building to transition towards a more data-intensive approach.

## 5.2 Reasons for optimism

For discussion purposes, let us assume that big data privacy concerns could be comprehensively addressed, that the skills and techniques could be put in place to make use of large datasets, and that useful data could be collected in the first instance. Could big data resolve long-standing unknowns in energy economy models? Could conditions of *data abundance* radically transform the way in which energy economy models are built, used, applied and interpreted, and change how decision makers think about energy and climate policy? We speculate that this could indeed be the case on all counts, although we make no claims about the likely speed with which these developments might unfold.

Economic modelling, and economics in general, has been called “an inexact and separate science” for decades, with an often fundamentally different outlook on issues such as truth, replicability, and validation to the natural sciences [140]. As previously discussed in Section 4.2, one of the longest running unresolved challenges in energy economics is the inability of researchers to harmonise their understanding of the energy system from both a *top-down* (monetary flow) and *bottom-up* (physical system) perspective. These are currently extremely difficult to reconcile together and one approach can sometimes give highly divergent insights from the other [141]. The IMACLIM model has perhaps come closest in fusing the two perspectives, but its use has not yet become common due to the challenges of co-calibrating the physical and monetary input-output matrix that underlays the model [142].

Better data on the relationships between technology diffusion, prices, and regulation (i.e. empirical Category 1 data, as discussed in Section 3.1) could significantly enhance our understanding of energy system dynamics, including the actual tendencies of the system to equilibrium and disequilibrium, its response to shocks, and other non-linearities. Understanding these relationships could give new insights into how decision makers might manipulate the system to address critical challenges such as reducing emissions [5], improving air quality [143], or providing modern energy services to all [144]. In future, decision makers might even be able to design both short and long term policies that precision target individual sub-sectors or systems rather than being limited to applying broad based “shotgun” interventions. Finally, following the principles of adaptive management pioneered in fisheries by Holling [145], and subsequently adopted by many others (e.g. [146]), policies could also potentially be designed so that data on their effects could be collected and assessed in real time in order to allow continuous policy assessment, fine-tuning or wholesale adjustment.

The use of energy models is already deeply intertwined with the practice of scenario analysis and the constructive application of the imagination to problem solving [60]. This is because forward-looking estimates of future system conditions will always contain elements of uncertainty, no matter how accurate the calibration of models to retrospective

empirical data (even *big data*). Unsurprisingly, the possibilities explored in models have been shown to influence beliefs about future possibilities [147]. Better computation in the 1970s and 1980s was central to spreading the idea that policymakers could outmanoeuvre threats to energy security through investing in energy efficiency and switching to alternative fuel sources instead of simply thinking of economic well-being in relation to oil use. In the same way, a step-change in data availability and quality might broaden policymakers' thinking on the solution space for climate mitigation, such as the ability to test changes in economic structure or focusing on the provision of energy services instead of fuels and equipment (e.g. thinking in terms of supplying mobility rather than cars).

### 5.3 Revolutionising the state-of-the-art

Flowing from the findings of this review, we propose a set of short, medium and long term recommendations for future experiments with energy economy modelling and big data.

**Near-term:** Our first, shorter term set of recommendations involves setting up experiments to test how big data can inform critical relationships used in energy economy models. At this stage there will be critical methodological developments and learning needed on both sides on issues such as on the data mining of existing databases to extract useful parameters for models, data curation, and the process of interchanging information between big data sets and energy economy modelling parameters. As a result of the need to prototype and standardise these kinds of working practices between big data experts and energy modelling practitioners, it is prudent to first start with a small set of tractable energy economy research questions. An excellent starting point could be to explore the use of big data to inform the behavioural choice algorithms of the type applied in transport mode choice models (e.g. [148]) and throughout behaviourally explicit whole system models such as the CIMS [43] modelling platform (i.e. derivative logit models), where behavioural components or preferences are added along with the standard capital, labour, energy and operating costs. There are several modelling teams operating globally that employ *bottom-up hybrid* models, and any one of them could begin the process of using big data to derive the parameters for their sector specific behavioural algorithms. This would involve a multi-step work program of assessing a prospective area of interest (e.g. urban transport mode choice), developing a big data set to work with, reformulating the mode choice algorithms in a *bottom-up hybrid* model to accommodate the big data, validating the model by testing how well it predicts history, and then beginning to use the model to explore the effects of both existing and prospective policy experiments.

**Mid-term:** In the medium term, a more ambitious project for an experienced energy modelling team would be to collaborate with experts in big data and machine learning, using their combined experience to completely reconfigure an existing *bottom-up hybrid* model or to construct a completely new energy model using state-of-the-art data fusion and integration methods [136]. This would be then used for deriving (1) the underlying multi-level structure of energy supply and demand from actual empirical use data (i.e. Category 1, as discussed in Section 3.1), and (2) the key behavioural determinants of energy use in the economy. In all likelihood the energy system structure would need to be initially seeded with a hypothetical system (e.g. parts of an existing model), which machine learning algorithms would then be able to use to sequentially test against real data, thus

revealing the underlying system. Early tests could begin with mode choice data for transport, due to the likely voluminous data that will be generated in this area (see Section 4.1), but researchers might also be able to use large volumes of aggregated smart meter data on end-use demand in households and firms if it is available.

**Long-term:** Finally, over the longer term, a more speculative set of research questions and programs can be imagined. As discussed earlier, energy modellers are arguably still working in a kind of “Henry Ford” era of modelling, where there is little customisation, and where research is often performed with general purpose, one-size-fits-all type frameworks, e.g. MARKAL/TIMES. As a result, energy models are currently extremely labour intensive to apply. Scientists are forced to build up large institutions behind general purpose models and then adapt them as best they can to meet the structure and data requirements of specific research and policy questions. Could a world of *data abundance* facilitate on-demand construction of more useful “bespoke” models, built from real world empirical data (Category 1) at the appropriate social, economic and technical level of detail for any given policy problem? Could it be possible to conceive of a data and modelling structure that could adapt dynamically to the research question of the moment? Could scientists invoke semi-autonomous agents to both curate and mine databases [149], and have them construct energy models on-the-fly according to user-specified parameters, such as geographic scale, sectors of interest and overall theoretical structure (e.g. a *top-down hybrid* CGE framework, or a *bottom-up hybrid* simulation framework)? The possibilities are intriguing, and we encourage the energy modelling community to think beyond the constraints imposed by current modelling paradigms and imagine the cutting edge possibilities that might exist in this space.

## 6.0 Conclusions

While other fields of scientific research are working quickly to take advantage of the possibilities offered by big data, energy economy modelling is lagging, both in theory and in practice. In this paper, we have identified several areas where big data sources could help to add more detail in crucial areas for existing models. Additionally, we believe the tantalising prospect of completely revolutionising energy modelling may now lie on the horizon. Big data combined with other next generation information technologies might resolve long standing barriers between engineering-derived and economics-focused views of the energy system, and in a way that both eliminates critical blind spots and creates new possibilities for dynamic, adaptive climate policymaking. We suggest that one specific area in the near term where this could occur is through dynamic parameterization of an existing *bottom-up hybrid* model with situation specific behavioural algorithms derived from big data. Another, in the longer term, is where big data, combined with existing theory, is allowed to shape the structure and behavioural dynamics of energy supply and demand models to more accurately represent these complex and situation specific systems.

While the human imagination and the use of scenario assumptions about the future will likely always be required for energy and climate policy analysis, big data could potentially reduce much of the reliance on data extrapolation and (fallible) expert judgement that exists in most models when characterising energy systems. For this potential promise to be fulfilled however, several developmental steps are required. First, issues surrounding data

collection costs, ownership and privacy must be addressed. Second, energy modellers must become familiar with the data collection and analysis tools of big data. Third, energy modellers must be willing to look dispassionately at the basic dynamic assumptions and datasets underpinning their models, and answer honestly whether increased availability of Category 1 and Category 2 data can transform them from being theoretically interesting *substitute* models to policy useful *surrogate* models. Finally, there must be a willingness in the energy modelling community to embrace fundamental changes in their modelling paradigms, and to embrace the new spring of *data abundance* after living through a long winter of *data scarcity*.

## 7.0 Acknowledgements

Funding support for Francis Li was provided by the UK Engineering and Physical Sciences Research Council (EPSRC) under the projects “Modelling the Political, Societal, and Regulatory Implementation of the UK Energy Decarbonisation Transition” [Grant EP/R002096/1] and the “Whole Systems Energy Modelling Consortium (wholeSEM)” [Grant EP/K039326/1]. Chris Bataille was supported by the French government under the framework of the programme “Investissements d’avenir”, managed by ANR (the French National Research Agency) under the reference ANR-10-LABX-01, and also received support that helped lead to this article under a wholeSEM Fellowship at University College London [Grant EP/K039326/1].

## 8.0 References

- [1] Bataille C, Waisman H, Colombier M, Segafredo L, Williams J, Jotzo F. The need for national deep decarbonization pathways for effective climate policy. *Clim Policy* 2016;16:57–26. doi:10.1080/14693062.2016.1173005.
- [2] Taylor PG, Upham P, McDowall W, Christopherson D. Energy model, boundary object and societal lens: 35 years of the MARKAL model in the UK. *Energy Res Soc Sci* 2014;4:32–41. doi:10.1016/j.erss.2014.08.007.
- [3] Clarke L, Jiang K, Akimoto K, Babiker M, Blanford G, Fisher-Vanden K, et al. Assessing Transformation Pathways. In: Edenhofer O, Pichs-Madruga R, Sokona Y, Farahani E, Kadner S, Seyboth K, et al., editors. *Clim. Chang. 2014 Mitig. Clim. Chang. Contrib. Work. Gr. III to Fifth Assess. Rep. Intergov. Panel Clim. Chang.*, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2014, p. 413–510.
- [4] Pye S, Li FGN, Price J, Fais B. Achieving net-zero emissions through the reframing of UK national targets in the post-Paris Agreement era. *Nat Energy* 2017;2:17–24. doi:10.1038/nenergy.2017.24.
- [5] UNFCCC. Adoption of the Paris Agreement: Proposal by the President: Draft decision -/CP.21. United Nations Framework Convention on Climate Change (UNFCCC): 2015.

- [6] Rogelj J, Popp A, Calvin K V., Luderer G, Emmerling J, Gernaat D, et al. Scenarios towards limiting global mean temperature increase below 1.5 °C. *Nat Clim Chang* 2018;8:325–32. doi:10.1038/s41558-018-0091-3.
- [7] Lempert R, Popper S, Bankes S. Confronting Surprise. *Soc Sci Comput Rev* 2002;20:420–40. doi:10.1177/089443902237320.
- [8] Mäki U. On the method of isolation in economics. *Pozn Stud Philos Sci Humanit* 1992;26:19–54.
- [9] Hourcade J-C, Jaccard M, Bataille C, Gherzi F. Hybrid Modeling: New Answers to Old Challenges Introduction to the Special Issue of The Energy Journal. *Energy J* 2006;SI2006. doi:10.5547/ISSN0195-6574-EJ-VolSI2006-NoSI2-1.
- [10] Bell G, Hey T, Szalay A. Beyond the Data Deluge. *Science (80- )* 2009;323:1297–8. doi:10.1126/science.1170411.
- [11] Hilbert M, Lopez P. The World’s Technological Capacity to Store, Communicate, and Compute Information. *Science (80- )* 2011;332:60–5. doi:10.1126/science.1200970.
- [12] Fosso Wamba S, Akter S, Edwards A, Chopin G, Gnanzou D. How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study.’ *Int J Prod Econ* 2015;165:234–46. doi:10.1016/j.ijpe.2014.12.031.
- [13] Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Heal Inf Sci Syst* 2014;2. doi:10.1186/2047-2501-2-3.
- [14] Erevelles S, Fukawa N, Swayne L. Big Data consumer analytics and the transformation of marketing. *J Bus Res* 2016;69:897–904. doi:10.1016/j.jbusres.2015.07.001.
- [15] Kacfeh Emani C, Cullot N, Nicolle C. Understandable Big Data: A survey. *Comput Sci Rev* 2015;17:70–81. doi:10.1016/j.cosrev.2015.05.002.
- [16] Fan S, Lau RYK, Zhao JL. Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. *Big Data Res* 2015;2:28–32. doi:10.1016/j.bdr.2015.02.006.
- [17] Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* 2007;22:338–42. doi:10.1096/fj.07-9492LSF.
- [18] Wang Z, Wang H, Zhao W, Cheng L. Energy optimization of parallel programs in a heterogeneous system by combining processor core-shutdown and dynamic voltage scaling. *Futur Gener Comput Syst* 2019;92:198–209. doi:10.1016/j.future.2018.09.039.
- [19] Di Salvo ALA, Agostinho F, Almeida CMVB, Giannetti BF. Can cloud computing be labeled as “green”? Insights under an environmental accounting perspective. *Renew Sustain Energy Rev* 2017;69:514–26.

- doi:10.1016/j.rser.2016.11.153.
- [20] Zhou K, Yang S. Understanding household energy consumption behavior: The contribution of energy big data analytics. *Renew Sustain Energy Rev* 2016;56:810–9. doi:10.1016/j.rser.2015.12.001.
  - [21] Linder L, Vionnet D, Bacher J-P, Hennebert J. Big Building Data - a Big Data Platform for Smart Buildings. *Energy Procedia* 2017;122:589–94.
  - [22] E.Edwards R, New J, E.Parker L, Cui B, Dong J. Constructing large scale surrogate models from big data and artificial intelligence. *Appl Energy* 2007;202:685–99. doi:10.1016/j.apenergy.2017.05.155.
  - [23] Reinhart CF, Cerezo Davila C. Urban building energy modeling – A review of a nascent field. *Build Environ* 2016;97:196–202. doi:10.1016/j.buildenv.2015.12.001.
  - [24] Batty M. Big data, smart cities and city planning. *Dialogues Hum Geogr* 2013;3:274–9. doi:10.1177/2043820613513390.
  - [25] Rathore MM, Ahmad A, Paul A, Rho S. Urban planning and building smart cities based on the Internet of Things using Big Data analytics. *Comput Networks* 2016;101:63–80. doi:10.1016/j.comnet.2015.12.023.
  - [26] Tao F, Cheng J, Qi Q, Zhang M, Zhang H, Sui F. Digital twin-driven product design, manufacturing and service with big data. *Int J Adv Manuf Technol* 2018;94:3563–76. doi:10.1007/s00170-017-0233-1.
  - [27] Chen C, Liu Y, Kumar M, Qin J. Energy Consumption Modelling Using Deep Learning Technique — A Case Study of EAF. *Procedia CIRP* 2018;72:1063–8. doi:10.1016/j.procir.2018.03.095.
  - [28] Liang YC, Lu X, Li WD, Wang S. Cyber Physical System and Big Data enabled energy efficient machining optimisation. *J Clean Prod* 2018;187:46–62. doi:10.1016/j.jclepro.2018.03.149.
  - [29] Alahakoon D, Yu X. Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey. *IEEE Trans Ind Informatics* 2016;12:425–36. doi:10.1109/TII.2015.2414355.
  - [30] Akhavan-Hejazi H, Mohsenian-Rad H. Power systems big data analytics: An assessment of paradigm shift barriers and prospects. *Energy Reports* 2018;4:91–100. doi:10.1016/j.egyr.2017.11.002.
  - [31] Faizrahneem M, Schlote A, Maggi L, Crisostomi E, Shorten R. A big-data model for multi-modal public transportation with application to macroscopic control and optimisation. *Int J Control* 2015;88:2354–68. doi:10.1080/00207179.2015.1043582.
  - [32] Kargashin PE, Prasolova AI, Novakovskiy BA, Rafikova YY, Gorbunov R V., Gorbunova TY. Data processing as a critical part of GIS based mapping of



- renewable energy perspectives. *MATEC Web Conf* 2018;178:09004.  
doi:10.1051/mateconf/201817809004.
- [33] Sarasti J. Evaluation of load shifting potential in Espoo district heating network using a model predictive approach. Aalto University, 2017.
- [34] Sivarajah U, Kamal MM, Irani Z, Weerakkody V. Critical analysis of Big Data challenges and analytical methods. *J Bus Res* 2017;70:263–86.  
doi:10.1016/j.jbusres.2016.08.001.
- [35] Philip Chen CL, Zhang C-Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Inf Sci (Ny)* 2014;275:314–47.  
doi:10.1016/j.ins.2014.01.015.
- [36] Ong K-L, De Silva D, Boo YL, Lim EH, Bodi F, Alahakoon D, et al. Big Data Applications in Engineering and Science. In: Yu S, Guo S, editors. *Big Data Concepts, Theor. Appl.*, Cham: Springer International Publishing; 2016, p. 315–51. doi:10.1007/978-3-319-27763-9\_9.
- [37] Labrinidis A, Jagadish H V. Challenges and opportunities with big data. *Proc VLDB Endow* 2012;5:2032–3. doi:10.14778/2367502.2367572.
- [38] Bello-Orgaz G, Jung JJ, Camacho D. Social big data: Recent achievements and new challenges. *Inf Fusion* 2016;28:45–59. doi:10.1016/j.inffus.2015.08.005.
- [39] Zhang Y, Zhao Y. Astronomy in the Big Data Era. *Data Sci J* 2015;14:11.  
doi:10.5334/dsj-2015-011.
- [40] Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. *Int J Inf Manage* 2015;35:137–44. doi:10.1016/j.ijinfomgt.2014.10.007.
- [41] Suppes P. Models of Data. In: Nagel E, Suppes P, Tarski A, editors. *Stud. Log. Found. Math.*, 1966, p. 252–61. doi:10.1016/S0049-237X(09)70592-0.
- [42] Li FGN, Pye S. Uncertainty, politics, and technology: Expert perceptions on energy transitions in the United Kingdom. *Energy Res Soc Sci* 2018;37:122–132.  
doi:10.1016/j.erss.2017.10.003.
- [43] Bataille C, Jaccard M, Nyboer J, Rivers N. Towards General Equilibrium in a Technology-Rich Model with Empirically Estimated Behavioral Parameters. *Energy J* 2006;SI2006. doi:10.5547/ISSN0195-6574-EJ-VolSI2006-NoSI2-5.
- [44] Frigg R, Hartmann S. Models in Science. In: Zalta EN, editor. *Stanford Encycl. Philos.*, Stanford, CA, USA: 2012.
- [45] Knight FH. *Risk, Uncertainty, and Profit*. New York City, NY, USA: Hart, Schaffner and Marx; 1921.
- [46] Wynne B. Uncertainty and environmental learning: Reconceiving science and policy in the preventive paradigm. *Glob Environ Chang* 1992;2:111–27.  
doi:10.1016/0959-3780(92)90017-2.

- [47] Stirling A. Keep it complex. *Nature* 2010;468:1029–31. doi:10.1038/4681029a.
- [48] Boyd D, Crawford K. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Commun Soc* 2012;15:662–79. doi:10.1080/1369118X.2012.678878.
- [49] Häggqvist S. A Model for Thought Experiments. *Can J Philos* 2009;39:55–76. doi:10.1353/cjp.0.0040.
- [50] Hughes RIG. Models and Representation. *Philos Sci* 1997;64:5325–36. doi:10.1086/392611.
- [51] Parker WS. Does matter really matter? Computer simulations, experiments, and materiality. *Synthese* 2009;169:483–96. doi:10.1007/s11229-008-9434-3.
- [52] Mäki U. Models are experiments, experiments are models. *J Econ Methodol* 2005;12:303–15. doi:10.1080/13501780500086255.
- [53] Robinson JB. Of maps and territories. *Technol Forecast Soc Change* 1992;42:147–64. doi:10.1016/0040-1625(92)90005-E.
- [54] Kriegler E, Riahi K, Petermann N, Bosetti V, Capros P, Vuuren DP van, et al. Assessing Pathways toward Ambitious Climate Targets at the Global and European Levels: A synthesis of results from the AMPERE project. 2014.
- [55] Pye S, Li FGN, Petersen A, Broad O, McDowall W, Price J, et al. Assessing qualitative and quantitative dimensions of uncertainty in energy modelling for policy support in the United Kingdom. *Energy Res Soc Sci* 2018;46:332–44. doi:10.1016/j.erss.2018.07.028.
- [56] Walker WE, Harremoës P, Rotmans J, van der Sluijs JP, van Asselt MBA, Janssen P, et al. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integr Assess* 2003;4:5–17. doi:10.1076/iaij.4.1.5.16466.
- [57] Lee J, Lapira E, Bagheri B, Kao H. Recent advances and trends in predictive manufacturing systems in big data environment. *Manuf Lett* 2013;1:38–41. doi:10.1016/j.mfglet.2013.09.005.
- [58] Mäki U. *Realistic Realism about Unrealistic Models*. vol. 1. Oxford University Press; 2009. doi:10.1093/oxfordhb/9780195189254.003.0004.
- [59] Godfrey-Smith P. The strategy of model-based science. *Biol Philos* 2006;21:725–40. doi:10.1007/s10539-006-9054-6.
- [60] Frigg R. Models and fiction. *Synthese* 2010;172:268. doi:10.1007/s11229-009-9505-0.
- [61] Bhosekar A, Ierapetritou M. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Comput Chem Eng* 2018;108:250–67. doi:10.1016/j.compchemeng.2017.09.017.

- [62] van der Hoog S. Surrogate Modelling in (and of) Agent-Based Models: A Prospectus. *Comput Econ* 2018;1–19. doi:10.1007/s10614-018-9802-0.
- [63] Mäki U. Models and the locus of their truth. *Synthese* 2011;180:47–63. doi:10.1007/s11229-009-9566-0.
- [64] Sugden R. Credible worlds: the status of theoretical models in economics. *J Econ Methodol* 2000;7:1–31. doi:10.1080/135017800362220.
- [65] Sugden R. Credible Worlds, Capacities and Mechanisms. *Erkenntnis* 2009;70:3–27. doi:10.1007/s10670-008-9134-x.
- [66] Cash DW, Clark WC, Alcock F, Dickson NM, Eckley N, Guston DH, et al. Knowledge systems for sustainable development. *Proc Natl Acad Sci* 2003;100:8086–91. doi:10.1073/pnas.1231332100.
- [67] Li FGN, Pye S, Strachan N. Regional winners and losers in future UK energy system transitions. *Energy Strateg Rev* 2016;13–14:11–31. doi:10.1016/j.esr.2016.08.002.
- [68] McGlade C, Ekins P. The geographical distribution of fossil fuels unused when limiting global warming to 2 °C. *Nature* 2015;517:187–90. doi:10.1038/nature14016.
- [69] Deane JP, Chiodi A, Gargiulo M, Ó Gallachóir BP. Soft-linking of a power systems model to an energy systems model. *Energy* 2012;42:303–12. doi:10.1016/j.energy.2012.03.052.
- [70] Ramachandran K, Turton H. Switzerland Energy Transition Scenarios – Development and Application of the Swiss TIMES Energy System Model (STEM). Paul Scherrer Institut (PSI): 2014.
- [71] Li P-H, Pye S. Assessing the benefits of demand-side flexibility in residential and transport sectors from an integrated energy systems perspective. *Appl Energy* 2018;228:965–79. doi:10.1016/j.apenergy.2018.06.153.
- [72] Fais B, Sabio N, Strachan N. The critical role of the industrial sector in reaching long-term emission reduction, energy efficiency and renewable targets. *Appl Energy* 2016;162:699–712. doi:10.1016/j.apenergy.2015.10.112.
- [73] Rausch S, Metcalf GE, Reilly JM. Distributional impacts of carbon pricing: A general equilibrium approach with micro-data for households. *Energy Econ* 2011;33:S20–33. doi:10.1016/j.eneco.2011.07.023.
- [74] Lamb WF, Steinberger JK, Bows-Larkin A, Peters GP, Roberts JT, Wood FR. Transitions in pathways of human development and carbon emissions. *Environ Res Lett* 2014;9:014011. doi:10.1088/1748-9326/9/1/014011.
- [75] Crespo Cuaresma J. Income projections for climate change research: A framework based on human capital dynamics. *Glob Environ Chang* 2017;42:226–36. doi:10.1016/j.gloenvcha.2015.02.012.

- [76] Vandyck T, Van Regemorter D. Distributional and regional economic impact of energy taxes in Belgium. *Energy Policy* 2014;72:190–203. doi:10.1016/j.enpol.2014.04.004.
- [77] Wilson C, Dowlatabadi H. Models of Decision Making and Residential Energy Use. *Annu Rev Environ Resour* 2007;32:169–203. doi:10.1146/annurev.energy.32.053006.141137.
- [78] Loulou R, Remne U, Kanudia A, Lehtila A, Goldstein G. Documentation for the TIMES Model. International Energy Agency Energy Technology Systems Analysis Programme (IEA-ETSAP): 2005.
- [79] McCollum DL, Wilson C, Pettifor H, Ramea K, Krey V, Riahi K, et al. Improving the behavioral realism of global integrated assessment models: An application to consumers' vehicle choices. *Transp Res Part D Transp Environ* 2017;55:322–42. doi:10.1016/j.trd.2016.04.003.
- [80] Li FGN. Actors behaving badly: Exploring the modelling of non-optimal behaviour in energy transitions. *Energy Strateg Rev* 2017;15:57–71. doi:10.1016/j.esr.2017.01.002.
- [81] Pye S, Bataille C. Improving deep decarbonization modelling capacity for developed and developing country contexts. *Clim Policy* 2016;16:S27–46. doi:10.1080/14693062.2016.1173004.
- [82] Koseleva N, Ropaite G. Big Data in Building Energy Efficiency: Understanding of Big Data and Main Challenges. *Procedia Eng* 2017;172:544–9. doi:10.1016/j.proeng.2017.02.064.
- [83] Schuelke-Leech B-A, Barry B, Muratori M, Yurkovich BJ. Big Data issues and opportunities for electric utilities. *Renew Sustain Energy Rev* 2015;52:937–47. doi:10.1016/j.rser.2015.07.128.
- [84] Kamargianni M, Li W, Matyas M, Schäfer A. A Critical Review of New Mobility Services for Urban Transport. *Transp Res Procedia* 2016;14:3294–303.
- [85] Shahrokni H, Levihn F, Brandt N. Big meter data analysis of the energy efficiency potential in Stockholm's building stock. *Energy Build* 2014;78:153–64. doi:10.1016/j.enbuild.2014.04.017.
- [86] Mathew PA, Dunn LN, Sohn MD, Mercado A, Custudio C, Walter T. Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. *Appl Energy* 2015;140:85–93. doi:10.1016/j.apenergy.2014.11.042.
- [87] Wang S, Zhang Q, Martin R V, Philip S, Liu F, Li M, et al. Satellite measurements oversee China's sulfur dioxide emission reductions from coal-fired power plants. *Environ Res Lett* 2015;10. doi:10.1088/1748-9326/10/11/114015.
- [88] Pearce W, Holmberg K, Hellsten I, Nerlich B. Climate Change on Twitter:

- Topics, Communities and Conversations about the 2013 IPCC Working Group 1 Report. *PLoS One* 2014;9. doi:10.1371/journal.pone.0094785.
- [89] Chen M, Mao S, Liu Y. Big Data: A Survey. *Mob Networks Appl* 2014;19:171–209. doi:10.1007/s11036-013-0489-0.
- [90] Zhou K, Fu C, Yang S. Big data driven smart energy management: From big data to big insights. *Renew Sustain Energy Rev* 2016;56:215–25. doi:10.1016/j.rser.2015.11.050.
- [91] Kavousian A, Rajagopal R, Fischer M. Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy* 2013;55:184–94. doi:10.1016/j.energy.2013.03.086.
- [92] Albert A, Rajagopal R. Smart Meter Driven Segmentation: What Your Consumption Says About You. *IEEE Trans Power Syst* 2013;28:4019–30. doi:10.1109/TPWRS.2013.2266122.
- [93] Roberts S. Demographics, energy and our homes. *Energy Policy* 2008;36:4630–2. doi:10.1016/j.enpol.2008.09.064.
- [94] Scarpa R, Willis K. Willingness-to-pay for renewable energy: Primary and discretionary choice of British households' for micro-generation technologies. *Energy Econ* 2010;32:129–36. doi:10.1016/j.eneco.2009.06.004.
- [95] Green J, Newman P. Citizen utilities: The emerging power paradigm. *Energy Policy* 2017;105:283–93.
- [96] Rylatt M, Gammon R, Boait P, Varga L, Allen P, Savill M, et al. CASCADE: An Agent Based Framework for Modeling The Dynamics of Smart Electricity Systems. *Emerg Complex Organ* 2013;15:1–13.
- [97] Szabó S, Moner-Girona M, Kougias I, Bailis R, Bódis K. Identification of advantageous electricity generation options in sub-Saharan Africa integrating existing resources. *Nat Energy* 2016;1. doi:10.1038/nenergy.2016.140.
- [98] Veefkind JP, Aben I, McMullan K, Förster H, De Vries J, Otter G, et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sens Environ* 2012;120:70–83. doi:10.1016/j.rse.2011.09.027.
- [99] Doll CNH, Pachauri S. Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery. *Energy Policy* 2010;38:5661–70. doi:10.1016/j.enpol.2010.05.014.
- [100] Ma J, Cheng JCP. Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology. *Appl Energy* 2016;183:182–92. doi:10.1016/j.apenergy.2016.08.079.

- [101] Evans S, Liddiard R, Steadman P. Modelling a whole building stock: domestic, non- domestic and mixed use. *Build Res Inf* 2017. doi:10.1080/09613218.2017.1410424.
- [102] DeFries R, Achard F, Brown S, Herold M, Murdiyarso D, Schlamadinger B, et al. Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. *Environ Sci Policy* 2007;10:385–94. doi:10.1016/j.envsci.2007.01.010.
- [103] González MC, Hidalgo CA, Barabási A-L. Understanding individual human mobility patterns. *Nature* 2008;453:779–782. doi:10.1038/nature06958.
- [104] Simini F, González MC, Maritan A, Barabási A-L. A universal model for mobility and migration patterns. *Nature* 2012;484:96–100. doi:10.1038/nature10856.
- [105] Van Alphen K, Van Voorst tot Voorst Q, P.Hekkert M, Smits REHM. Societal acceptance of carbon capture and storage technologies. *Energy Policy* 2007;35:4368–80. doi:10.1016/j.enpol.2007.03.006.
- [106] Baker SR, Bloom N, Davis SJ. Measuring Economic Policy Uncertainty. *Q J Econ* 2016;131:1593–1636. doi:10.1093/qje/qjw024.
- [107] Blanford GJ, Merrick JH, Bistline JET, Young DT. Simulating Annual Variation in Load, Wind, and Solar by Representative Hour Selection. *Energy J* 2018;39. doi:10.5547/01956574.39.3.gbla.
- [108] Arias MB, Bae S. Electric vehicle charging demand forecasting model based on big data technologies. *Appl Energy* 2016;183:327–39. doi:10.1016/j.apenergy.2016.08.080.
- [109] Weiss M, Helfenstein A, Mattern F, Staake T. Leveraging smart meter data to recognize home appliances. 2012 IEEE Int. Conf. Pervasive Comput. Commun., Lugano, Switzerland: 2012. doi:10.1109/PerCom.2012.6199866.
- [110] UN DESA. World Urbanization Prospects, The 2011 Revision. New York, USA: 2011.
- [111] Buffat R, Froemelt A, Heeren N, Raubal M, Hellweg S. Big data GIS analysis for novel approaches in building stock modelling. *Appl Energy* 2017;208:277–90. doi:10.1016/j.apenergy.2017.10.041.
- [112] Paltsev S, Reilly JM, Jacoby HD, Morris JF. The cost of climate policy in the United States. *Energy Econ* 2009;31:S235–43. doi:10.1016/j.eneco.2009.06.005.
- [113] Tuladhar SD, Yuan M, Bernstein P, Montgomery WD, Smith A. A top-down bottom-up modeling approach to climate change policy analysis. *Energy Econ* 2009;31:S223–34. doi:10.1016/j.eneco.2009.07.007.
- [114] Fishbone LG, Abilock H. Markal, a linear-programming model for energy systems analysis: Technical description of the BNL version. *Int J Energy Res* 1981;5:353–75. doi:10.1002/er.4440050406.

- [115] Böhringer C, Rutherford TF. Combining bottom-up and top-down. *Energy Econ* 2008;30:574–96. doi:10.1016/j.eneco.2007.03.004.
- [116] Sue Wing I. The synthesis of bottom-up and top-down approaches to climate policy modeling: Electric power technology detail in a social accounting framework. *Energy Econ* 2008;30:547–73. doi:10.1016/j.eneco.2006.06.004.
- [117] Li FGN, Strachan N. Take me to your leader: Using socio-technical energy transitions (STET) modelling to explore the role of actors in decarbonisation pathways. *Energy Res Soc Sci* 2019;51:67–81. doi:10.1016/j.erss.2018.12.010.
- [118] Jaccard MK, Nyboer J, Bataille C, Sadownik B. Modeling the Cost of Climate Policy: Distinguishing Between Alternative Cost Definitions and Long-Run Cost Dynamics. *Energy J* 2003;24. doi:10.5547/ISSN0195-6574-EJ-Vol24-No1-3.
- [119] Waisman H, Guivarch C, Grazi F, Hourcade JC. The Imaclim-R model: infrastructures, technical inertia and the costs of low carbon futures under imperfect foresight. *Clim Change* 2012;114:101–20. doi:10.1007/s10584-011-0387-z.
- [120] Hasselmann K, Kovalevsky D V. Simulating animal spirits in actor-based environmental models. *Environ Model Softw* 2013;44:10–24. doi:10.1016/j.envsoft.2012.04.007.
- [121] Bennear LS, Stavins RN. Second-best theory and the use of multiple policy instruments. *Environ Resour Econ* 2007;37:111–29. doi:10.1007/s10640-007-9110-y.
- [122] Li FGN, Strachan N. Modelling energy transitions for climate targets under landscape and actor inertia. *Environ Innov Soc Transitions* 2017;24:106–29. doi:10.1016/j.eist.2016.08.002.
- [123] Rivers N, Jaccard M. Combining Top-Down and Bottom-Up Approaches to Energy-Economy Modeling Using Discrete Choice Methods. *Energy J* 2005;26. doi:10.5547/ISSN0195-6574-EJ-Vol26-No1-4.
- [124] Axsen J, Mountain DC, Jaccard M. Combining stated and revealed choice research to simulate the neighbor effect: The case of hybrid-electric vehicles. *Resour Energy Econ* 2009;31:221–38. doi:10.1016/j.reseneeco.2009.02.001.
- [125] Jaccard M, Dennis M. Estimating home energy decision parameters for a hybrid energy-economy policy model. *Environ Model Assess* 2006;11:91–100. doi:10.1007/s10666-005-9036-0.
- [126] Michelsen CC, Madlener R. Homeowners' preferences for adopting innovative residential heating systems: A discrete choice analysis for Germany. *Energy Econ* 2012;34:1271–83. doi:10.1016/j.eneco.2012.06.009.
- [127] Brand C, Cluzel C, Anable J. Modeling the uptake of plug-in vehicles in a heterogeneous car market using a consumer segmentation approach. *Transp Res Part A Policy Pract* 2017;97:121–36. doi:10.1016/j.tra.2017.01.017.

- [128] Eastin MS, Brinson NH, Doorey A, Wilcox G. Living in a big data world: Predicting mobile commerce activity through privacy concerns. *Comput Human Behav* 2016;58:214–20. doi:10.1016/j.chb.2015.12.050.
- [129] Lu R, Zhu H, Liu X, Liu J, Shao J. Toward efficient and privacy-preserving computing in big data era. *IEEE Netw* 2014;28:46–50. doi:10.1109/MNET.2014.6863131.
- [130] Colombo P, Ferrari E. Privacy Aware Access Control for Big Data: A Research Roadmap. *Big Data Res* 2015;2:145–54. doi:10.1016/j.bdr.2015.08.001.
- [131] Tene O, Polonetsky J. Privacy in the Age of Big Data: A Time For Big Decisions. *Stanford Law Rev* 2012;64:63–9.
- [132] Jifa G, Lingling Z. Data, DIKW, Big Data and Data Science. *Procedia Comput Sci* 2014;31:814–21. doi:10.1016/j.procs.2014.05.332.
- [133] Hand DJ, Mannila H, Smyth P. *Principles of Data Mining*. MIT Press; 2001.
- [134] Mohri M, Rostamizadeh A, Talwalkar A. *Foundations of Machine Learning*. 2018.
- [135] Keim D, Qu H, Ma K-L. Big-Data Visualization. *IEEE Comput Graph Appl* 2013;33:20–1. doi:10.1109/MCG.2013.54.
- [136] Zheng Y. Methodologies for Cross-Domain Data Fusion: An Overview. *IEEE Trans Big Data* 2015;1:16–34. doi:10.1109/TBDDATA.2015.2465959.
- [137] Golfarelli M, Maio D, Rizzi S. The Dimensional Fact Mode: A Conceptual Model for Data Warehouses. *Int J Coop Inf Syst* 1998;07:215–47. doi:10.1142/S0218843098000118.
- [138] Wang L. Heterogeneous Data and Big Data Analytics. *Autom Control Inf Sci* 2017;3:8–15. doi:10.12691/acis-3-1-3.
- [139] Gobble MM. Resources: Big Data: The Next Big Thing in Innovation. *Res Manag* 2013;56:64–7. doi:10.5437/08956308X5601005.
- [140] Hausman DM. *The inexact and separate science of economics*. Cambridge, UK: Cambridge University Press; 1992. doi:10.1017/CBO9780511752032.
- [141] Dai H, Mischke P, Xie X, Xie Y, Masui T. Closing the gap? Top-down versus bottom-up projections of China's regional energy use and CO<sub>2</sub> emissions. *Appl Energy* 2016;162:1355–73. doi:10.1016/j.apenergy.2015.06.069.
- [142] Sassi O, Crassous R, Hourcade JC, Gitz V, Waisman H, Guivarch C. IMACLIM-R: a modelling framework to simulate sustainable development pathways. *Int J Glob Environ Issues* 2010;10:5. doi:10.1504/IJGENVI.2010.030566.
- [143] Lott MC, Pye S, Dodds PE. Quantifying the co-impacts of energy sector decarbonisation on outdoor air pollution in the United Kingdom. *Energy Policy* 2017;101:42–51. doi:10.1016/j.enpol.2016.11.028.



- [144] Fuso Nerini F, Tomei J, To LS, Bisaga I, Parikh P, Black M, et al. Mapping synergies and trade-offs between energy and the Sustainable Development Goals. *Nat Energy* 2018;3:10–5. doi:10.1038/s41560-017-0036-5.
- [145] Holling CS. *Adaptive environmental assessment and management*. Chinchester, USA: John Wiley & Sons; 1978.
- [146] Mathy S, Criqui P, Knoop K, Fishedick M, Samadi S. Uncertainty management and the dynamic adjustment of deep decarbonization pathways. *Clim Policy* 2016;16:S47–62. doi:10.1080/14693062.2016.1179618.
- [147] Borup M, Brown N, Konrad K, Van Lente H. The sociology of expectations in science and technology. *Technol Anal Strateg Manag* 2006;18:285–98. doi:10.1080/09537320600777002.
- [148] Daly HE, Ramea K, Chiodi A, Yeh S, Gargiulo M, Gallachóir BÓ. Incorporating travel behaviour and travel time into TIMES energy system models. *Appl Energy* 2014;135:429–39. doi:10.1016/j.apenergy.2014.08.051.
- [149] Cao L, Gorodetsky V, Mitkas PA. Agent Mining: The Synergy of Agents and Data Mining. *IEEE Intell Syst* 2009;24:64–72. doi:10.1109/MIS.2009.45.
- [150] Lempert RJ, Popper SW, Bankes SC. *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis*. RAND report MR-1626. RAND Corporation, The RAND Pardee Center, Santa Monica, CA, USA: RAND Corporation; 2003.