

Characterising the Evolution of Energy System Models Using Model Archaeology

Paul E. Dodds · Ilkka Keppo · Neil Strachan

Received: 21 June 2013 / Accepted: 22 May 2014

© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract In common with other types of complex models, energy system models have opaque structures, making it difficult to understand both changes between model versions and the extent of changes described in research papers. In this paper, we develop the principle of model archaeology as a formal method to quantitatively examine the balance and evolution of energy system models, through the ex post analysis of both model inputs and outputs using a series of metrics. These metrics help us to understand how models are developed and used and are a powerful tool for effectively targeting future model improvements. The usefulness of model archaeology is demonstrated in a case study examining the UK MARKAL model. We show how model development has been influenced by the interests of the UK government and the research projects funding model development. Despite these influences, there is clear evidence of a strategy to balance model complexity and accuracy when changes are made. We identify some important long-term trends including higher technology capital costs in subsequent model versions. Finally, we discuss how model archaeology can improve the transparency of research model studies.

Keywords Model archaeology · Energy system model · UK MARKAL · Transparency · Ex post analysis

1 Introduction

1.1 The Importance of Transparency in Model Evolution

Similar to complex modelling across a range of disciplines, energy system models need to be as clear and transparent as

possible to ensure quality assurance for users and replicability for practitioners [1]. Historically, modellers' efforts in this regard have been in terms of comparable documentation [2], model comparison exercises [3] and a very limited attempt at ex post evaluation of modelling results [4].

Model transparency and repeatability are even more relevant for energy system models as these technology-rich, economic optimisation models, such as MARKAL/TIMES [5], MESSAGE [6] and OSeMOSYS [7], have become critical tools for informing policy and business decisions in low-carbon energy technologies in many countries [e.g. 8, 9]. One issue for many consumers of the outputs of these models is that the complex structures, containing thousands of resources, technologies, commodities and energy demands (the 'reference energy system' or RES), tend to make the models opaque to outsiders.

Model understanding and transparency would be difficult enough if energy system models were static; however, these complex modelling tools are frequently changed within a cycle of model development, policy use and application to decision-making [10]. As a result, such model changes, which cause important changes in the model outputs, are frequently described in many different places. Yet, understanding the importance of model changes is very important for the research and the policy-making communities. Model changes are often summarised in the literature, but full descriptions would be too verbose to be provided in many cases. Even if a full description is provided, it is often difficult for an outsider to assess the extent and importance of the changes. The new model might be better than the old model or might simply be different, depending on the reasons for the changes. Updates might, for example, facilitate the exploration of new research questions, improve the scope of the model or update the model data. There is also a very real danger that research papers present results from a research model that no longer exists after the paper is published. Inadequate model documentation

P. E. Dodds (✉) · I. Keppo · N. Strachan
UCL Energy Institute, 14 Upper Woburn Place,
London WC1H 0NN, UK
e-mail: p.dodds@ucl.ac.uk

reduces the transparency and repeatability of an experiment to the extent that DeCarolus et al. [1] argue that energy system models may no longer be classed as scientific models.

Ex post analysis of model evolution can help us to understand the drivers of model development, including the strategies used to keep the model up-to-date and the influences of interested parties such as researchers and policy makers on the development process. In this study, we propose a methodology that we call ‘model archaeology’ to quantitatively examine the evolution of energy system models through the ex post analyses of both model inputs and outputs. This approach allows us to understand how such models change over time and can help us to more easily explain why different versions produce different results and what the different results may imply for the interpretation of the results. We believe that our quantitative approach complements and improves upon the more traditional qualitative descriptions of model changes that one normally finds in the literature. Model archaeology also complements the limited literature [e.g. 4, 11] on ex post analysis of model outputs (although these studies assess the skill of different model versions against actual developments of the energy system rather than comparing the outputs of different versions of the same model, as we do in this paper).

We know of no studies that combine ex post analyses of inputs and outputs. In this paper, we demonstrate the benefits of model archaeology in a case study of the UK MARKAL model. We hope that other researchers will use model archaeology to objectively assess the extent of model changes in research studies and to better understand how models have evolved, particularly in terms of the model balance between detail and complexity (both between and within sectors of the economy), so that the technique can be refined in the future.

1.2 Process of Model Evolution

Complex energy system models are constantly evolving. This is due to a range of interacting drivers including the requirements of policy makers, the availability of new data, the focuses of new projects, movements to new software platforms, research team staff turnover and the emergence of new ‘hot topics’ in the interdisciplinary field of energy modelling. Modellers can choose to follow any of a number of formal or informal processes to change their models. One variation between models is whether there is a demarcation between production and research versions of a model. In this paper, we refer to the principal version of a model as the ‘production’ version, and we define any variations of the model that are produced from the production version during research projects as ‘research’ versions.

In our experience, many, but not all, of the developments from research projects are included in the following production model with the expectation that they will improve the model. Changes often increase the model complexity but do

not necessarily make it better; in fact, if there is no discernible improvement in the model accuracy, then some argue that the model is more opaque, more difficult to understand and therefore an inferior model. Moreover, model development through research projects is likely to be piecemeal rather than strategic and could cause the model to become unbalanced.

Other modellers choose to separate the production and research versions of their model. For these models, research versions are produced to answer specific research questions, and a smaller proportion of the changes are likely to be incorporated into the production version, which is likely to be more transparent and less complicated than models that incorporate all research findings. Variations in model outputs between versions are likely to be less significant under this more conservative regime, which might make the model more valuable to some stakeholders (e.g. policy makers looking for a model that produces robust results over several versions) but could lead to divergence between production and research version results over time. Such a conservative approach could unconsciously cause a reluctance to make widespread changes to the model in response to new information and could potentially contribute to poor policy decisions in a process termed negative learning [12].

Finding sufficient time to keep a model up-to-date can be difficult, particularly for modellers who rely on short-term funding; it can be difficult to gain research funding purely for strategic model updating, particularly because measuring improvements in models is so difficult and often not valued [10]. Research funding is likely to focus on model application and on the sectors of most interest to other academics and policy makers; so, research-led model development is likely to be skewed towards particular sectors. We show in this paper how model archaeology can be used to identify such long-term trends.

1.3 Balancing Model Detail and Complexity

An important challenge for modellers is to balance model detail and complexity. Model improvement often takes the form of increasing the detail of model sectors, which increases the model complexity. One reason for this is to gain greater insights into decarbonisation pathways within particular sectors [e.g. 13, 14]. Increasing model complexity is also viewed as a strategy by some modellers to improve the model performance¹ and to reduce the structural uncertainty. For technology-rich models such as those based on the MARKAL/TIMES model generator, improvements tend to target the RES or model topology rather than the underlying

¹ It is not possible to assess whether model performance has improved for most energy system models as they project future scenarios. Moreover, adding complexity does not necessarily improve the model performance: A black box approach can be more accurate than an explicit description that is unable to capture some key characteristics of a phenomenon.

paradigm and equations, despite the latter being a comparable source of structural uncertainty (see the uncertainty matrix in [15]). For example, studies have shown that the optimal take-up of energy efficiency improvements varies greatly depending on the model paradigm and boundaries [16, 17].

The disadvantage of increasing the complexity of a model, for example by increasing the size of the energy system structure, is that it tends to reduce the transparency and repeatability of any model-based experiments [1]. More complex models are more difficult to maintain and are more likely to contain erroneous data (which is also then more difficult for the modeller to find and correct). For these reasons, some authors contend that models should be as simple as possible within the bounds of the research question [18, 19].

Model archaeology can quantify changes to the model complexity over time and so can assist modellers in understanding whether there is an appropriate balance between complexity and detail. Moreover, the model archaeology metrics can also be used to compare the complexity of models using similar paradigms, which could be particularly useful for multi-model studies such as those performed by the Energy Modeling Forum (EMF).

1.4 Model Boundaries

The choice of model boundaries depends on the research question and can greatly affect the results. For example, many energy system models examine only CO₂ emissions, but an EMF model comparison study concluded that expanding model boundaries to include non-CO₂ greenhouse gases reduced the total cost of mitigation [20]. Yet, if non-CO₂ gases are added to a model, then it is necessary to expand the scope of the model beyond the energy system in order to include mitigation options for non-energy processes that produce greenhouse gas emissions, which will increase the model complexity.

Models have structural as well as topological boundaries. The model equations are difficult to change in some energy system model generators such as MARKAL/TIMES; so, structural changes are generally limited to model variants (e.g. elastic demand response rather than fixed energy service demands [21] or myopic or stochastic variants of the model), new versions of the model generator or new modelling systems with the energy system model linked to another model. In contrast, the MESSAGE model code is designed with flexibility in mind [22], and the OSEMOSS model is specifically designed to be easily changed, but at the costs of limiting its size and complexity [7]. Since model boundaries can change in subsequent model versions, they should be considered in a model archaeology analysis.

1.5 Outline of This Study

In this paper, we formalise the application of model archaeology to energy system models in Section 2. We then illustrate the usefulness of model archaeology in a case study examining the UK MARKAL model. In Section 3, we examine how the model inputs have changed over time, and we use this in Section 4 to identify stages of model evolution and the influence of policy and research interests. In Section 5, we investigate how the UK MARKAL outputs vary between versions and how these variations are related to changes in the model inputs. We look for relationships between inputs and outputs in Section 6; then, we conclude by reflecting on the value of model archaeology in Section 7.

2 Model Archaeology

Model archaeology is different to more traditional model comparisons both because it assesses both ex post model inputs and outputs and uniquely characterises the evolution of a complex model using a series of qualitative and quantitative metrics. It differs from the ex post analyses documented in the literature [e.g. 11], in that comparisons are made against other versions of the model, not against actual developments of the energy system. The method therefore does not attempt to assess whether past projections were accurate² but focuses on the development of the tool itself. In this section, we formalise the application of model archaeology to technology-rich energy system models by defining a series of metrics to characterise changes in the model design and the model outputs.

2.1 Design Metrics

Since the aim of model archaeology is to characterise the development of the model as fully as possible, the choice of design metrics should depend on the design and purpose of the model. For energy system models, we propose that metrics should cover the following aspects:

1. Model paradigm and equations
2. Spatial and temporal dimensions
3. Energy system structure (model topology)
4. Modelled system constraints
5. Parameter data

² Whether ‘accurate’ projections are even the goal of an exercise is another question: Most of the large energy system models are not used as forecasting tools, and they rather provide normative, optimised scenarios, in which many real implementation bottlenecks are ignored (e.g. uncertainty, heterogeneity of decision makers and market imperfections)

We collect these statistics for major versions of the UK MARKAL model to assess the model evolution. We can then understand how the model has changed by comparing these statistics against funded projects, publications and other outputs.

2.1.1 Model Equations and Variants

Since few changes to the model equations are likely in policy-focused models (Section 1.4), these are best described qualitatively rather than quantitatively. Some models are designed with the purpose of examining structural uncertainty by generating many variations of the equations (e.g. [23, 24]). The metrics we are proposing are not appropriate for such models.

The model boundaries might evolve over time. Where these involve changes in the paradigm, for example to add a climate module to a global energy system model, then these changes should be reported as a paradigm change. Other changes that only affect the energy system structure should be reported in that section.

2.1.2 Spatial and Temporal Dimensions

The spatial dimension of an energy system model is normally defined by the geographical area covered by the model and the number of internal and external regions.³ The temporal dimension is defined both in terms of the length of each time period (normally in years) and the number of time slices in each year (which can be seasonal, intra-day, etc.) for representing commodities such as electricity that are not easily stored.

2.1.3 Energy System Structure

The RES structure is likely to vary between different model versions. RESs are generally composed of many linked technologies, commodities, materials, demands and emission counters, and quantitative metrics are therefore used for giving a first-order estimate of the magnitude of the changes.

The first step is to characterise the overall RES in terms of the overall complexity, technology diversity and model boundaries. We assess the overall complexity by counting the number of technologies and commodities in the RES and the number of links between these technologies and commodities. We evaluate the technology diversity by counting the number of non-vintaged,⁴ non-

dummy⁵ technologies. We assess the model energy system boundaries qualitatively.

The second step is to characterise sectoral model development by disaggregating the RES by sector and evaluating the technology diversity in each sector. At this stage, it is also useful to assess the model balance according to the purpose of the model. Since most models are designed to provide insights into future energy system configurations and greenhouse gas reductions, useful metrics are a comparison of the technology diversity with the nominal energy service demands and the emissions from each sector. Since the technology diversity is likely to be sensitive to the heterogeneity of the sector (for example, the UK non-residential building stock is much more diverse than the residential building stock but has a much lower energy demand), these statistics are most useful for comparing different model versions rather than for judging the balance of a single model.

Many models have several inter-linked regions, each with a separate RES, and these could be analysed independently or as a whole (in many energy system models, for example TIAM-UCL [25] and MESSAGE [22], each region has the same RES structure). It might also be beneficial to examine the number of links between regions, for example to assess the number of regionally traded commodities.

2.1.4 Modelled System Constraints

System constraints are used for a variety of reasons, for example to put limits on the growth or use of groups of technologies, to implement complicated technologies or to take account of non-modelled phenomena. System constraints are different to constraints that affect only single technologies or resources, which we analyse in “Parameter Data”.

Many modellers find that the number of system constraints tends to increase over time and this can unnecessarily restrict model performance. For this reason, we evaluate the number of system constraints in each model sector. Another approach is to categorise each system constraint according to its purpose and to sum each type of constraint for comparison with other models. For example, categories might include system constraints required for correct model operation, government policies, technology limitations, consumer preferences and constraints to fix model weaknesses.

2.1.5 Parameter Data

Parameter data are the data of each technology, commodity, demand and emission. In a mature model, one might expect the parameter data to be changed more often than the RES structure. We assess parametric changes for energy system

³ The full energy system of each internal region is simulated endogenously in the model, and there are cross-border commodity flows where appropriate. External regions are outside of the model boundary and provide only imports and exports of commodities to the internal regions.

⁴ Each vintage of a technology has the same properties (operating costs, energy efficiency) throughout its lifetime, even if improved versions of the technology become available. In some models, such as MARKAL, it is necessary to define each vintage as a separate technology, while in other models, such as TIMES, they can all be represented using a single technology.

⁵ Dummy technologies are those used for accounting or modelling purposes and do not represent an actual technology.

models by categorising similar types of data for each technology and evaluating whether a change has been made in each category. The categories we use are cost (including resource, capital and O&M costs), commodity flows (including inputs and outputs) and constraints (investment, capacity or activity constraints/bounds on particular technologies). This approach also enables us to find out whether some categories are changed more than others.

2.2 Output Metrics

The choice of output metrics depends to some extent on the model design and, in particular, the research questions that the model is designed to answer. This means that the choice of output metrics is less well defined than the choice of input metrics.

It is necessary to carefully define a suitable range of scenarios for analysis in all model versions. These scenarios should be consistent with the principal research questions.

For all technology-rich models, we would expect comparing the primary and final energy consumption in each sector to be useful. For those models that are used to examine decarbonisation pathways, the greenhouse gas emissions from each sector should also be examined. Other potentially useful metrics are energy flows, plant capacities and investment levels in sectors of particular interest. It might also be enlightening to examine how particular commodities are used in different model versions, for example electricity or bioenergy products, to understand the impact of changes to the model inputs.

2.3 Relationships Between Input and Output Metrics

Model archaeology can be a much more powerful tool if we can quantitatively compare how much model results change as a function of changes to the inputs. We might, for example, examine changes between model versions or compare changes over all of the versions in different sectors. Such metrics can show us whether output changes are more sensitive to some types of input changes than others (e.g. the energy system structure rather than parameter data) and whether the quantity or quality of the changes tends to have a greater effect on the results. We can also use these metrics to compare the relationships between inputs and outputs in different energy system models.

3 Application of Model Archaeology Input Metrics to UK MARKAL

In this section, we examine the evolution of the UK MARKAL energy system model using model archaeology by comparing the inputs of model versions over a 6-year period.

UK MARKAL was originally developed by the UK government for the 2003 Energy White Paper [26] and subsequently adopted by Ricardo-AEA on privatisation. The current version of UK MARKAL was subsequently developed at University College London (UCL). It portrays the entire UK energy system from imports and domestic production of fuel resources, through fuel processing and supply, representation of infrastructures, conversion of fuels to secondary energy carriers (including electricity and heat), end-use technologies and energy service demands of the entire economy [27, 28].

3.1 Input Metrics

We apply the model archaeology metrics to the six production versions of UK MARKAL that are highlighted in Fig. 1.

3.1.1 Model Equations and Variants

UK MARKAL has always been run using the ANSWER interface, and the underlying equations have never been changed within the MARKAL source code. However, several MARKAL variants have been developed over the last 8 years, including a hybrid macro research version of the model [29]. Elastic energy service demands have been used in the production version since v3.17 [21], and a stochastic study recently examined the impact of relaxing the perfect foresight paradigm [30].

3.1.2 Spatial and Temporal Dimensions

UK MARKAL has a single internal region that represents the UK in all production versions. The exceptions are the following: (i) the spatial hydrogen research version of the model, which disaggregates the UK into nine demand regions, six supply points and a set of 200 infrastructure development options for the purpose of constructing hydrogen pipelines; and (ii) the two-region model that includes Scotland as an individual region.

UK MARKAL uses 5-year time periods from 2000 until 2070 in earlier versions and 2050 from v3.25 onwards. The model was first developed at a time when time slices were restricted to three seasons (summer, winter and intermediate) and two intra-day slices (day, night). These six time slices are used in all versions except for a temporal research model which included four seasons and five intra-day time slices. The temporal version was created by the same project as the spatial hydrogen version.

3.1.3 Reference Energy System Structure

The evolution of the production versions of UK MARKAL in terms of the number of technologies/commodities/demands and the number of links between these nodes is shown in

Fig. 1 Timeline of production and research versions of the UK MARKAL model. The production versions analysed in this paper are highlighted. All production versions since v3.17 have used the elastic demand variant of the model. Hybrid macro and stochastic variants have also been produced from production versions by research projects. Key publications for the research versions are cited in Section 3.2, and other model publications are cited in Section 4

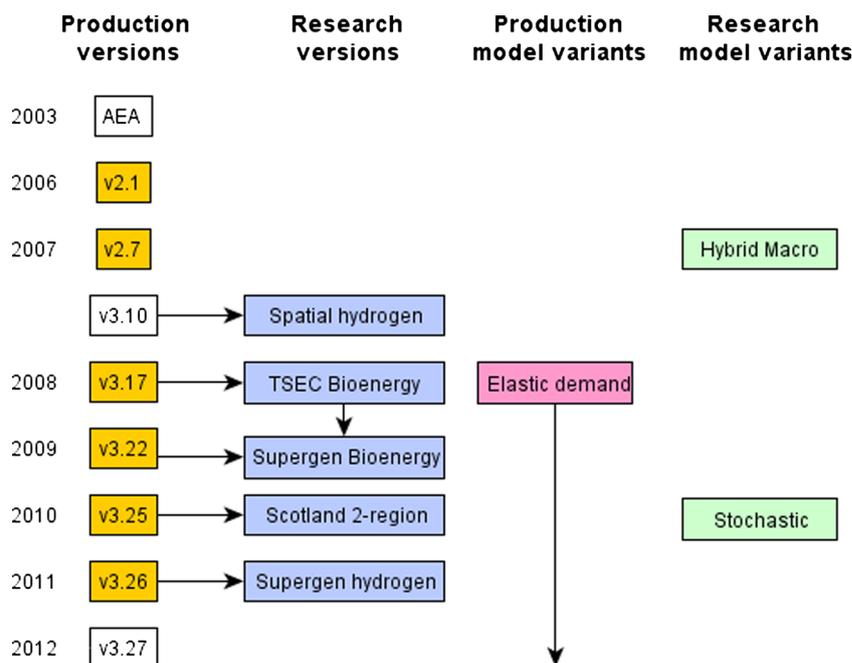


Fig. 2. The initial UCL version of the model was virtually a new model compared to the AEA version from 2003, as the number of technologies and links almost doubled. Since then, the total number technologies and links has increased slightly over time, but the graph shows that much of this increase is due to new dummy technologies being introduced for accounting or modelling purposes (e.g. to count the amount of fuel consumed in a subsector, for example for motorcycles within the transport sector). Over the 5 years of development shown here, the structure of the model technologies has been reasonably static, and model development has principally concentrated on improving technology data; when new technologies have been added, others have often been removed to avoid unnecessarily increasing the model complexity.

The UK MARKAL energy system boundaries have only been changed once. From v3.25, the RES was expanded to include the consumption of petroleum fuels as process feedstocks (e.g. bitumen production from oil). New energy service demands were defined for this change, which was implemented to enable the emissions from these processes to be counted for the first time.

We summarise the sectoral model development and balance in Table 1. The total number of diverse technologies increased by 16 % between v2.1 and v3.26, but the changes varied greatly between sectors, with the technology diversity reducing in the process and industry sectors. Almost 90 % of the v2.1 technologies are still present in v3.26, 5 years later. Table 1 also compares the technology diversity in v3.26 with UK CO₂ emissions and nominal energy service demands in each sector. The model looks well balanced in terms of the demands, which reflects that energy consumption statistics

rather than emission statistics were used in the design and calibration of UK MARKAL. At first sight, the model looks less well balanced in terms of CO₂ emissions, but process and service, the two sectors with high ratios, are more heterogeneous than other sectors; so, their relatively high numbers of diverse technologies are warranted.

The apparent stability in the UK MARKAL model does not mean that the structure of the model has received little attention apart from occasional additions. On the contrary, Table 2 shows that significant changes have taken place in several sectors, in particular for a major review that was completed by

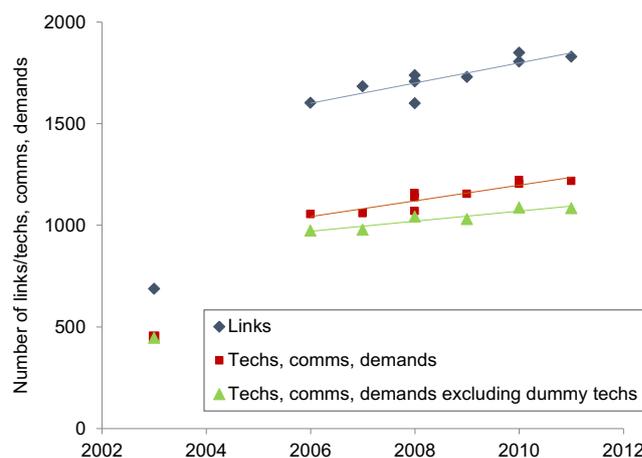


Fig. 2 Time series of the number of technologies, commodities and demands in the UK MARKAL model and the number of links between these nodes. Where technologies have several vintages, only the first vintage is counted. The number of links includes links to dummy technologies, but these are excluded from the green triangle series. The 2003 AEA version of the model is excluded from the trend lines

Table 1 UK MARKAL technology diversity and model balance. The first columns show the number of diverse (non-vintage, non-dummy) technologies in each sector in v2.1 and v3.26 of UK MARKAL. The final two columns show the technology diversity in terms of the CO₂ emissions for each UK sector in the year 2010 and the equivalent energy demand (which is estimated from the energy service demands using the most efficient demand technology available to the model in the year 2050. The process and electricity CO₂ emissions are produced during centralised electricity generation or fuel conversion, respectively

Sector	Technologies		Fraction from v2.1 in v3.26	Tech diversity/ Mte CO ₂ emissions	Tech diversity/ TJ equivalent demand
	v2.1	v3.26			
Resources	117	130	97 %	–	–
Electricity	94	140	81 %	0.9	–
Process	114	111	78 %	2.8	–
Residential	92	123	93 %	1.4	66.3
Service	36	52	100 %	2.8	53.3
Industry	98	93	94 %	1.3	46.3
Transport	110	117	84 %	1.0	65.2
Total	661	766	89 %	1.5	57.5

v3.17. Many technologies were removed during this review in order to reduce model complexity although some of those in the residential sector were later reinstated in v3.22 as a result of further experimentation. While many changes have been made in the electricity, residential and transport sectors, the service sector has only been changed twice (to add technologies), and the industrial sector only received some minor changes in the major review in v3.17. The total model rate of RES structural change (combining technology additions and deletions) is 12 %/year relative to v2.1.⁶

Table 2 shows that model improvement is not necessarily a continuous process. For UK MARKAL, the model has been reviewed and revised every few years and used for studies in between these reviews. Such an approach has only been possible because of the availability of long-term funding for model development; if UK MARKAL had relied on the smaller one-off research grants, then much less development would have been possible and updates would have been more sporadic and less strategic.

3.1.4 Modelled System Constraints

The number of user-defined constraints in UK MARKAL has risen steadily over time as shown in Fig. 3. The apparent linear nature of this graph does not wholly represent the development of constraints in the model, which is similar to the RES development described above. Table 3 shows that the majority of the constraints affect the electricity, residential and transport

⁶ The rate of RES structural change excludes changes to technology parameters (e.g. costs, technical characteristics, etc.), which are separately assessed in Section 3.1.5.

sectors and that there has been a substantial number of additions and deletions, particularly in the major review in v3.17 when almost half of the constraints were changed. The total model rate of constraint change (combining additions and deletions) is 25 %/year.

3.1.5 Parameter Data

We categorise technology parameter changes into costs, flows and bounds, and we count the numbers of technologies with any change in these categories. Figure 4 shows the results for the UK MARKAL model. Cost parameter changes are most common in all revisions. The total model rate of parameter data changes (combining costs, flows and bounds) is 16 %/year, which is similar to the rate of model RES structural change.

Parameter data changes are broken down by sector in Table 4. The resource, electricity and residential sectors have received sustained attention with many changes in most versions. In contrast, few data changes have been implemented in the process, industry and service sectors (although the service sector has expanded substantially over time).

For those technologies *that are present in all model versions*, investment costs in the electricity sector have been increased for 39 % of technologies and decreased for only 8 %. While a similar trend has occurred in the transport sector (34 % increase, 19 % decrease), residential technology costs have been reduced (10 % increase, 23 % decrease). The cost of importing and mining resources has increased in 54 % of cases and decreased in only 4 %. Changes in the process efficiencies have similarly been focused on the electricity and transport sectors; in the electricity sector, process efficiencies have generally been reduced (28 %, compared to 7 % with increases), while a similar proportion has been increased and reduced in the transport sector (41 % decreased and 39 % increased). Most residential heat process efficiencies have been increased in later model versions, but few of these technologies were present in UK MARKAL v2.1.

Overall, the representation of the future energy system has become more pessimistic over time, with a trend towards higher costs and lower energy efficiencies. The macro variant of v2.7 was used to estimate the impact of decarbonisation on the UK economy, and implementing the most recent costs in v2.7 would increase these costs, although these increases would be offset by the addition of elastic demands and new mitigation options since v2.7.

Once technologies have been added to the RES, it is very rare that new cost parameters (e.g. operating costs) have been added in subsequent versions of the model. Parameter changes have tended to only alter existing data. Of those technologies with capital costs, only 80 % have fixed operating costs and

Table 2 Change in the number of diverse technologies over time in four sectors of UK MARKAL. “+” indicates additions and “-” indicates deletions

Version	Year	Electricity		Process		Residential		Transport	
		+	-	+	-	+	-	+	-
v2.7	2007	0 %	-0%	0 %	-0 %	0 %	-0 %	0 %	-0 %
v3.17	2008	40 %	-15%	17 %	-22 %	37 %	-82 %	19 %	-16 %
v3.22	2009	2 %	-2%	2 %	-1 %	135 %	-4 %	0 %	-0 %
v3.25	2010	10 %	-1%	3 %	-1 %	0 %	-0 %	4 %	-0 %
v3.26	2011	9 %	-1%	0 %	-0 %	4 %	-0 %	1 %	-2 %

only 20 % have variable operating costs; so, this is perhaps an area for future model improvement.

Energy service demands significantly influence model results and are particularly uncertain in the future, yet have been only rarely changed in UK MARKAL since v2.1. Some transport and industry demands were changed in v3.17 and some transport demands again in v3.25. All of the other demands remained constant between v2.1 and v3.26.

3.1.6 Overall Technology RES and Parameter Changes

It is instructive to finish this analysis of UK MARKAL by understanding the overall combined changes between the first of the current versions of the model, v2.1, and the most recent version analysed here, v3.26. Table 5 shows that most of the residential and electricity technologies and half of the resource and transport technologies from v2.1 have been altered in subsequent versions, but that few of the process, service and industry technologies have been changed. Overall, 58 % of the technologies in v3.26 have been added or changed in the 5 years since v2.1.

3.2 Research Versions of UK MARKAL

The relatively stable structure of the production versions of UK MARKAL is in marked contrast to the research versions. Table 6 shows how the research versions differ from the production versions from which they were derived in terms of the number of links and technologies/commodities/demands. All of the research versions are more complex than the production versions, as a result of adding additional

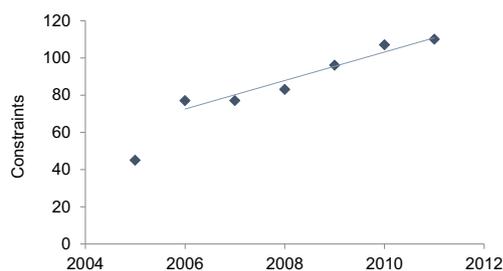


Fig. 3 Time series of the number of user-defined constraints in the UK MARKAL model. The AEA version of the model in 2005 is excluded from the trend line

regions, disaggregating sectors or simply adding detail by expanding the existing structure. The key elements and insights from the research versions have then been distilled into the next production version, following a strategy that avoids implementing wholesale research version changes in the production versions to ensure that the model does not become unnecessarily complicated.

4 The Evolution of the UK MARKAL Model

Model development is often a haphazard process [1], as we have demonstrated for UK MARKAL. We can use the ex post analysis of model inputs to identify four broad stages of development of UK MARKAL and to understand the drivers behind model development.

4.1 Stage 1: Initial Development

The modelling team received the AEA version of UK MARKAL in 2005 and initially planned only an extension, but v2.1 that emerged following extensive development was so radically different that it should be considered a new model. The aim at this early stage was to characterise the model [38]

Table 3 Number of constraints in each sector in UK MARKAL v3.26 and the total number of additions and deletions in each sector as the model has developed

	Constraints in v3.26	Total number of additions and deletions				
		v2.7	v3.17	v3.22	v3.25	v3.26
Resource	3	0	0	0	0	0
Electricity	24	0	11	1	10	3
Process	3	0	2	0	3	0
Residential	23	0	13	14	4	0
Service	6	0	1	0	2	0
Industry	9	0	0	0	2	0
Transport	42	0	23	0	6	0
Total	110	0	50	15	27	3

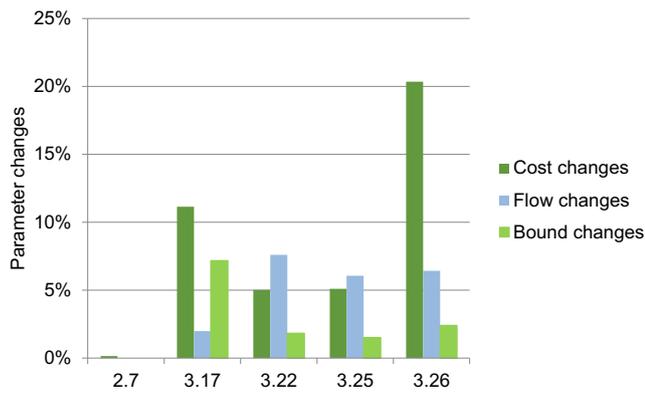


Fig. 4 Change in technology costs, flows and bounds between versions of the UK MARKAL model. Technologies that are added or removed from the RES are not included in these statistics

and to use it to identify UK decarbonisation pathways [39]. It had an important role in policy analyses for the Energy White Paper [40] and the Climate Change Bill [41]. A hybrid macro version of UK MARKAL was also created at this stage using v2.7 that was of particular interest to the UK government [29]. At the end of this stage, UK MARKAL had been used to support four major policy analyses [42].

4.2 Stage 2: Experimentation and Incremental Improvement

The second stage was marked by experimentation on the new model. Having tested the model and used it for initial studies, there was a drive to test the limits of the model by producing both spatial [31] and temporal [43] versions. These projects were supported by funding from a UK sustainable hydrogen energy project and from the UK Department for Transport, so led to substantial improvements to the transport sector in v3.17 for both technologies (Tables 3 and 5) and constraints (Table 3).

Model development was also supported by long-term funding from the UK Energy Research Centre, and an important new feature in v3.17, in 2008, was the addition of long-

run elastic demands [27]. Residential heat was reviewed [44], and the electricity sector received attention in most of these studies and was updated again. There has consistently been more interest in the decarbonisation of electricity generation than in any other sector in the UK, in both policy and research circles, and these have led to frequent reappraisals of technology data [e.g. 45, 46]. The high level of data availability has underpinned regular parameter updates in all model versions (Table 4).

The final important development in this stage was the production of a comprehensive manual [28] that greatly improved the transparency of the model.

4.3 Stage 3: Reflection

Having built, used and tested the limits of the model, the next stage could be characterised as a period of reflection. There was a consideration of how model scenarios should be constructed in the future [47]. The role of energy system models for UK policy analysis and research, and the appropriate level of funding for modelling activities, was questioned [10]. Meanwhile, projects began to examine model uncertainty by using stochastic decision-making to examine the impact of relaxing the perfect foresight paradigm in v3.25 [30] and through other methods.

The UK Climate Change Act 2008 [48] created the Committee on Climate Change (CCC) to assess the progress of the UK government towards reducing UK greenhouse gas emissions by 80 % in 2050. This led to several UK MARKAL studies examining decarbonisation pathways, including a study by UCL for the CCC [33] that led to the creation of v3.25 and a study by AEA for the Department of Energy and Climate Change [49] that created v3.26. The former included changes across the model (Table 4), including the addition of many new constraints (Table 3), while the latter concentrated on cost changes in the sectors of most interest to the government: electricity, residential, transport and resources.

4.4 Stage 4: Maturity and Reimagining

UK MARKAL has now reached a level of maturity. The uncertainty studies continue, and there is a continuing research interest in the benefits of expanding the model; for example, recent studies have disaggregated the transport and residential sectors [36, 50] and have examined the future of the UK gas networks [51, 13]. Efforts are being made to concentrate development in areas that have previously received comparatively less attention in the last 5 years; for example, the industry sector is now being completely revised in a major project. Particular weaknesses of the model framework, such as the representations of infrastructure and behaviour, are being studied in UK MARKAL for the first time.

Table 4 Fraction of technologies with data changes in each sector of each version of UK MARKAL

	Total number of parameter changes				
	v2.7	v3.17	v3.22	v3.25	v3.26
Resource	0 %	34 %	0 %	23 %	40 %
Electricity	18 %	19 %	32 %	19 %	39 %
Process	1 %	6 %	0 %	5 %	0 %
Residential	11 %	27 %	21 %	4 %	30 %
Service	56 %	0 %	0 %	0 %	0 %
Industry	0 %	0 %	1 %	3 %	0 %
Transport	0 %	11 %	0 %	35 %	25 %
Total	7 %	15 %	9 %	15 %	23 %

Table 5 Measures of UK MARKAL model changes, the fraction of technologies in v2.1 that have been subsequently changed and the fraction of technologies in v3.26 that have been changed or added since v2.1

Sector	Number of technologies		Fraction of v2.1 techs subsequently changed	Techs in v3.26 added or changed since v2.1
	v2.1	v3.26		
Resources	117	130	51 %	57 %
Electricity	134	177	75 %	83 %
Process	123	122	15 %	32 %
Residential	98	131	99 %	99 %
Service	40	56	58 %	71 %
Industry	133	128	3 %	4 %
Transport	540	511	30 %	45 %
Total	1,185	1255	47 %	55 %

In the longer term, there is a reimagining and renewal of the model. A new model (UKTM-UCL), based on the TIMES framework, is being produced to replace UK MARKAL. The broad design follows UK MARKAL, but it includes many new features, such as tracking emissions of several greenhouse gases rather than just CO₂ and including non-energy sources of emissions and mitigation measures. It includes the most recent advances from research models and addresses some of the weaknesses of UK MARKAL, for example by much improving the use of time slices throughout the model and by greatly improving the representation of energy storage. With the advent of UKTM-UCL, it is likely that these stages of development will repeat over the coming years.

4.5 Influence of Research Interests on UK MARKAL Development

The subject areas of the 22 journal papers and 4 major reports that use UK MARKAL are summarised in Table 7. Almost all

Table 6 Difference in links and technologies/commodities/demands between the research versions and the production versions of UK MARKAL

Research version	Production version	Links	Techs/comms/demands	Reference
Hybrid macro	v2.1	–	–	[29]
Spatial hydrogen	v3.10	+470 %	+267 %	[31]
TSEC bioenergy	v3.17	+2 %	+2 %	[14]
Supergen bioenergy	v3.22	+5 %	+8 %	[32]
Stochastic	v3.25	–	–	[33, 30]
Scotland 2-region	v3.25	+6 %	+9 %	[34]
SH process	v3.26	+18 %	+27 %	[35]
SH residential	v3.26	+11 %	+16 %	[36]
SH transport	v3.26	+54 %	+82 %	[37]

SH sustainable hydrogen models

Table 7 Summary of the journal papers and major reports using UK MARKAL. ‘Energy system’ refers to papers that examine the whole energy system. ‘Production’ and ‘Research’ refer to whether the studies used a production or research version of UK MARKAL

	Production	Research
Journal papers		
Energy system	7	1
Hydrogen	1	4
Bioenergy	0	2
Electricity	1	1
Residential	2	0
Transport	2	0
CCS	1	0
Major reports		
Energy system	4	0

of the papers that examine the whole energy system use the production version of the model (the Scotland 2-region research model is the exception). The presence of so many energy system journal papers is perhaps surprising; however, four of these examine macroeconomic and stochastic model variants while two others consider scenario construction methodologies. The number of research versions examining hydrogen and bioenergy reflect the specific project funding in these sectors. The electricity, residential and transport sectors are each the specific subject of two journal papers as well as forming the main focus of most of the energy system publications, which explains why they receive more attention in UK MARKAL than the other sectors.

While we can identify links between publications and model development, we cannot explain from this data why different parameters tend to receive attention in different sectors (e.g. investment costs for electricity and process efficiencies for residential heat). The availability of good-quality data is a key determinant that also affects model development, and this critically depends on the interests of stakeholders who fund the data collection. For example, early electricity sector decarbonisation has been identified as a key step towards meeting CO₂ targets by all versions of UK MARKAL [52], and the UK government has had a particular strong focus on the electricity sector since the 2003 Energy White Paper [4]; so, the government regularly commissions reports of the costs of electricity generation [e.g. 45, 46], and these reports are used to update the model. Conversely, the residential sector predominantly uses natural gas, and lower carbon fuels are not likely to be adopted for decades; so, there is much less data collection that could support model improvement (although initiatives such as the Renewable Heat Incentive [53] could provide more data in the future).

5 Application of Model Archaeology Output Metrics to UK MARKAL

In this section, we examine changes in the UK MARKAL outputs from each of the six versions examined in Section 3 and link these where possible to model development.

5.1 Methodology

The UK MARKAL model was originally developed to identify lowest-cost UK decarbonisation pathways, and this is still the primary purpose of the model today. We therefore examine the outputs using two scenarios based around long-term decarbonisation targets:

1. No CO₂ constraint: a base case with no limit on CO₂ emissions.
2. With CO₂ constraint: A linear stepwise reduction in CO₂ emissions from 2020 to 2050 that in total reduces emissions by 80 % in 2050 compared to 1990 emissions, in line with the requirements of the UK Climate Change Act [48]. For the 3.xx versions of UK MARKAL, we additionally examine the impact of including elastic demand responses to the price increases caused by moving to a low-carbon economy. The elastic demand reference prices for each version are taken from the ‘No CO₂ constraint’ scenario.

We analyse these scenarios in each of the six model versions. The final model period varies between 2050 and 2070 in different versions of the model, but we run each version only until 2050 to prevent the results from being affected by the model time horizon.

5.2 Results

The primary energy consumption in 2050 is shown for each scenario in each model version in Fig. 5. The variations between versions in the no CO₂ constraint scenario are small. For the scenario with a CO₂ constraint, nuclear electricity dominates in v2.1, but an increase in the cost of uranium enrichment in v2.7 causes much nuclear to be replaced by offshore wind generation. Biomass becomes increasingly important from v3.17 as a result of new biomass technologies introduced by a research project. The variations after v3.17 are smaller than between previous versions. Implementing elastic demand causes primary energy consumption to reduce by around 20 % in all four versions.

Final energy consumption is closely linked to energy service demand in all of the scenarios; so, there is little variation between sectors, except for transport, in any of the model versions (Fig. 6). Demand reductions in the elastic demand cases reduce final energy demand. The other particularly

notable trend is the addition of non-energy fuel consumption in v3.25, which is caused by the change in the model boundaries to include the non-energy use of petroleum-derived fuels.

5.2.1 Supply Sectors: Electricity Generation and Bioenergy

Electricity generation in each model version is shown in Fig. 7. For the no CO₂ constraint scenario, coal is the dominant feedstock in all versions, and the total generation varies little, despite this sector being one of the most often altered. There is much more variation in the scenarios with a CO₂ constraint. Nuclear and offshore wind dominate in v2.1 and v2.7. Coal CCS is deployed from v2.7 but is replaced when coal-biomass co-firing CCS is introduced to the model in v3.22. The variability in generation technologies between versions reflects the frequent changes to the electricity sector but is unlikely to continue in the future, as new constraints in v3.25 limit investment in each type of generation technology to around 2.5 GW/year. The model is likely to always pick a mixed portfolio of technologies under these constraints.

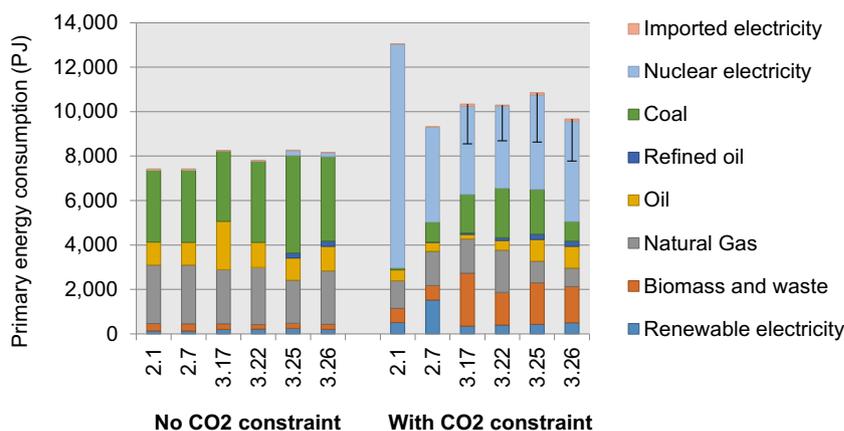
The reduction in total output from v3.17 reflects the increase in overall low-carbon electricity generation technology costs, which reduce the competitiveness of electricity against alternatives. This trend is somewhat reversed in v3.22 and v3.25 by the new option of atmospheric carbon sequestration through co-firing CCS with biomass; this technology enables the model to implement lower emission reductions in other sectors where emission cuts are more expensive so makes the electricity sector more competitive.

Figure 8 shows the consumption of bio-products in each model version. Bioenergy is only competitive in scenarios with a CO₂ constraint. Although numerous additional bioenergy technology routes are introduced into the process sector in v3.17, these are not responsible for the large increase in transport consumption of biofuels in that version. This increase is primarily driven by the increase in electricity costs; in v2.1 and v2.7, electricity is used to produce hydrogen through electrolysis, but this process becomes less competitive in v3.17, and ethanol vehicles, which use technologies from v2.1, become widespread instead. From v3.22, new co-firing CCS electricity technologies enable the model to sequester atmospheric carbon emissions, and biomass consumption switches to the electricity sector. A parameter change to increase the efficiency of biomass boilers in v3.26 causes biomass consumption to switch again to the residential sector and leads to the reduction in electricity generation shown in Fig. 7 for this version.

5.2.2 Demand Sectors: Transport, Residential and Industry

Transport sector fuel consumption is shown in Fig. 9 for each scenario. We assume that there will be only small cost

Fig. 5 Primary energy consumption in 2050 in each model version. Nuclear and renewable electricity generation are represented on the graph using the physical energy content method [54], following the International Energy Agency and the UK national statistics methodology [55]. The error bars show the reduction in energy consumption in the 'Elastic demand' scenario compared to the 'With CO₂ constraint' scenario



differences between internal combustion, battery and fuel cell drivetrains by 2050; so, the choice of transport fuel is sensitive to small price differentials and to changes in the other sectors. There are two principal decarbonisation options for transport: (i) hydrogen and electricity, which both generally depend on the price of electricity since hydrogen is produced by small-scale electrolysis in most cases with a CO₂ constraint; and (ii) biofuels, whose competitiveness is sensitive to the alternative uses of biomass in the energy system as shown in Fig. 8. This means that the switch from hydrogen in v2.1 and v2.7 to biofuels in v3.17 and back towards hydrogen by v3.26 is primarily related to model changes outside of the transport sector. Changes within the transport sector have less impact; for example, v3.22 has almost no changes yet is quite different to v3.17, while the many transport sector changes in v3.26 have little impact. Energy service demands are important; demand increases for buses, LGVs and HGVs contribute to the peaks in v3.17 and v3.22, but these demands are subsequently reduced again in v3.25. Introducing elastic demand

only slightly reduces transport demand in the scenario with a CO₂ constraint because fuel is only a small part of the total cost of ownership of a vehicle.

Residential heat generation in each scenario is shown in Fig. 10. Electric boilers dominate in v2.1 and v2.7 but become less competitive in v3.17 when the electricity price increases. Following the introduction of new biomass technologies in v3.17, an increase in gas prices in v3.25 induces a switch from natural gas boilers to wood-fuelled district heating. A substantial increase in the process efficiency of pellet boilers from 54 to 85 % in v3.26 stimulates a switch from district heating to biomass boilers. The introduction of pellet boilers is a good example of how a single important parameter change can profoundly affect the trajectories of several sectors of the economy.

Figure 11 shows the industrial fuel consumption in each scenario. There is a similar mix of fuels in all model versions, which reflects the constrained nature of

Fig. 6 Final energy consumption in 2050 in each model version. Non-energy use refers to the use of petroleum fuels as feedstock materials in industrial processes

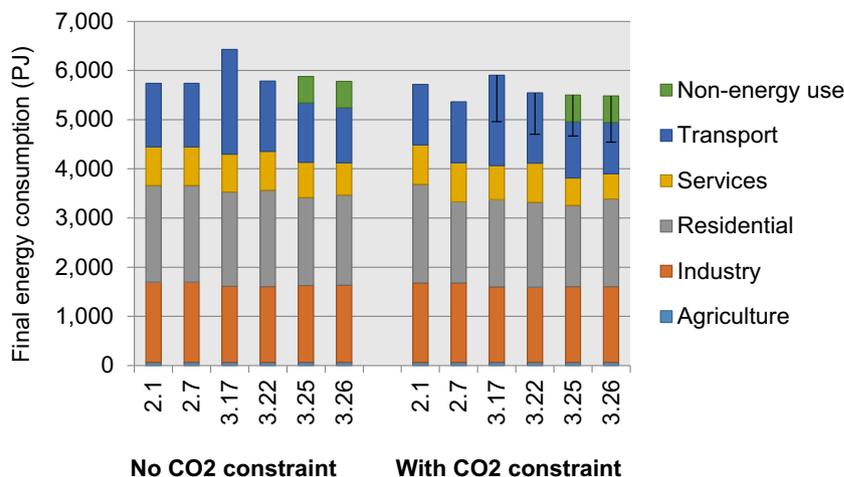
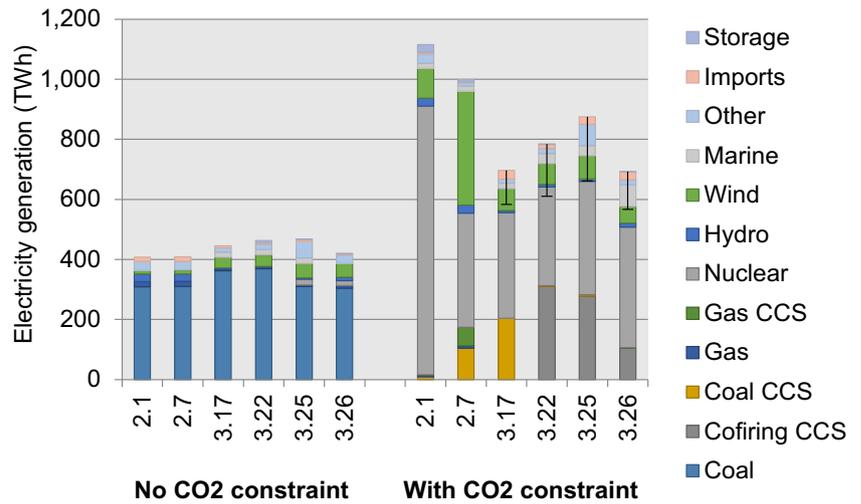


Fig. 7 Electricity generation in 2050 in each model version



the industrial sector in the model and the lack of changes to the sector since v2.1. The main difference between versions is the electricity consumption, which depends on the relative price of electricity to other fuels and is affected by changes elsewhere in the model.

5.2.3 Emissions

The CO₂ emissions from each sector are shown in Fig. 12. Only the scenarios with a CO₂ constraint are shown; so, the total emissions are 120 MtCO₂ in each scenario. Industrial sector emissions are relatively constant in each version, but the emissions from the electricity and residential sectors vary substantially. The introduction of co-firing CCS technologies in v3.22 enables the electricity sector to produce negative emissions, reducing the need for emission cuts in other sectors. If biomass CCS technologies were introduced to the model, then much higher negative emissions could be

achieved, and this would greatly influence technology choices in other sectors.

An important change from v3.25 is the change in the model boundaries to include the process feedstock consumption of petroleum fuels, which cause a substantial increase in CO₂ emissions (marked Other Emissions on Fig. 12). Since the model does not have abatement options for these emissions, it must make greater emission cuts in other sectors to compensate. The boundary change was introduced to improve the calibration of the model to UK energy and CO₂ emission statistics and would benefit from the addition of new abatement options in future versions of the model.

Figure 12 also shows the marginal CO₂ price in 2050 for the scenarios with and without elastic demand. Without elastic demand, this varies from £100/t to £150/t; it reduces with the introduction of co-firing in v3.22 but increases again when the model boundaries are changed in v3.25. The marginal price is

Fig. 8 Consumption of bioenergy products in each sector in 2050 in each model version

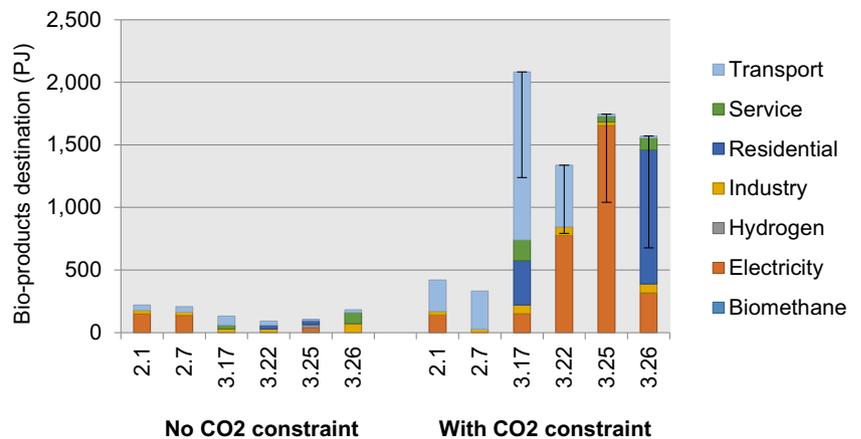
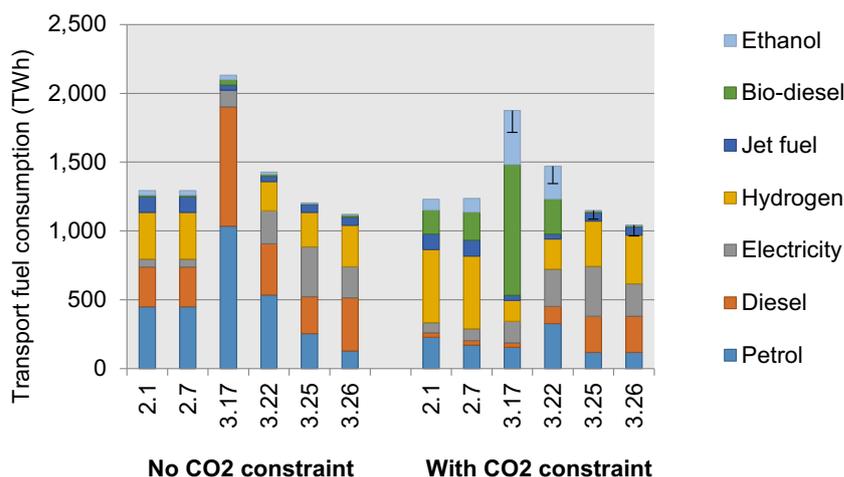


Fig. 9 Transport fuel consumption in 2050 in each model version



lower and more stable in the scenarios with elastic demand modelling.

6 Relationships Between Input and Output Metrics

We complete our model archaeology analysis of UK MARKAL by quantitatively relating changes in the model inputs to variations in the outputs, both within individual sectors and between model versions. Quantifying the magnitude of output variations in particular is not a trivial task. We suggest two methods in this section, and we hope to examine other methods in a future paper. We also hope to use such indices to quantitatively compare the evolution of other energy system models with UK MARKAL in the future.

6.1 Impact of Model Changes Within Sectors

It is useful to know if result variations within each sector are correlated with changes to the sectoral input data. In Sections 3.1.3 and 3.1.5, we examined the overall rate of change of model inputs per development

year by summing the number of changes (additions and deletions) to the energy system structure and to the system constraints and by counting the number of technology parameter data changes across all of the versions. For this investigation, we calculate the rate of change of inputs in each sector per year of model development across all model versions.

For our output index, we first calculate the coefficient of variation across all of the versions for each graph series. We then calculate the weighted mean of these coefficients of variation, with the weightings calculated using the mean of each series across all versions:

$$\text{Output index} = \frac{\sum_{x=1}^n \left(\frac{\sigma_x}{\bar{x}} \right) \bar{x}}{\sum_{x=1}^n \bar{x}}$$

where σ is the standard deviation and \bar{x} is the mean of series x across all model versions.

Table 8 compares the changes in the model inputs and outputs across all of the versions, in four sectors, for both

Fig. 10 Residential fuel consumption in 2050 in each model version

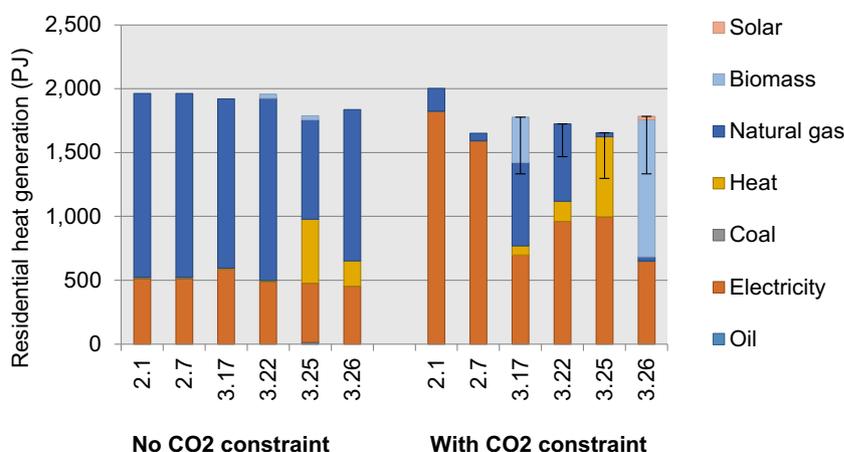
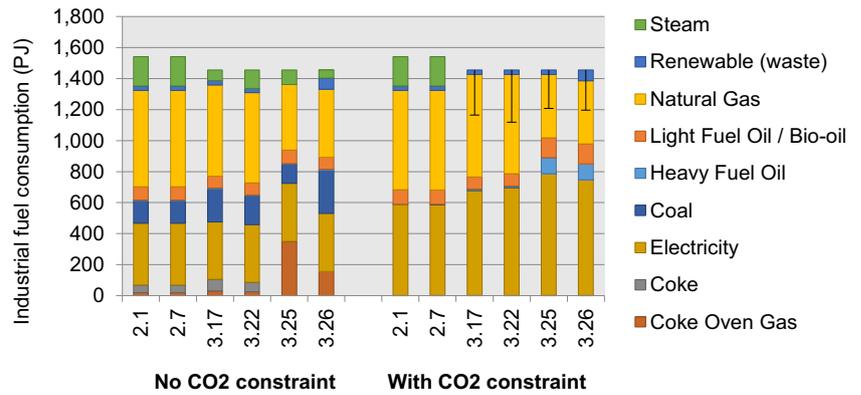


Fig. 11 Industrial fuel consumption in 2050 in each model version



scenarios (with and without a CO₂ constraint). Output variations are much lower for the cases with no CO₂ constraint and are not well correlated with changes to the inputs. In contrast, output variations in cases with a CO₂ constraint are well correlated for the three sectors with high numbers of input changes. The fourth sector, industry, has smaller but not insignificant output variations compared to the other sectors, showing that all of the sectors are affected to some extent by input changes in other sectors.

6.2 Model Changes Between Versions

It is necessary to use a different measure to calculate the rate of model input changes between versions. We instead calculate the number of changes in the model structure, the number of system constraints and the number of parameter changes for each version, relative to the previous version. We exclude the residential technologies that were removed in v3.17 and reinstated in v3.22 from the statistics.

We examine three model outputs: primary energy consumption, final energy consumption and CO₂ sectoral emissions. Since the model is primarily designed to identify decarbonisation pathways, we examine only the scenario with a CO₂ constraint. We use a different methodology for the model outputs than that in Section 6.1. For each graph series in Figs. 5, 6 and 12, we calculate the fractional change since the previous version. To ensure that the metric is not skewed

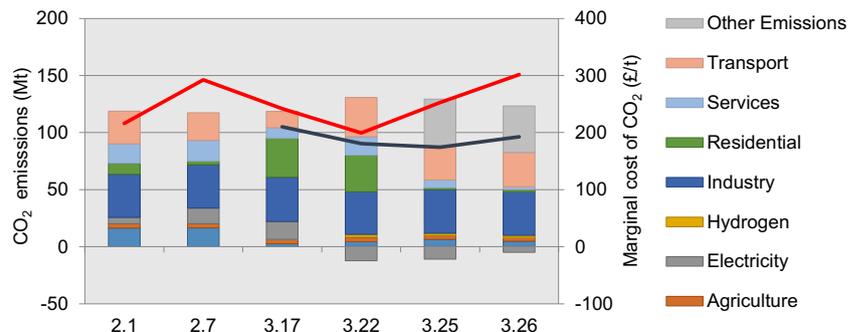
by negative CO₂ emissions in the denominator, we calculate the absolute (i.e. positive) change for each series. We avoid skewing the results through large fractional but small absolute changes by firstly assuming a maximum fractional change of 100 % between any two versions for each series and by secondly calculating the weighted mean of these percentage changes, with the weightings calculated using the mean of each series across all versions. These operations can be summarised as follows:

$$\text{Output index} = \frac{\sum_{x=1}^n \min\left(\text{abs}\left(\frac{x_v - x_{v-1}}{x_{v-1}}\right), 1\right) \bar{x}}{\sum_{x=1}^n \bar{x}}$$

where v is the model version, x is the graph series (e.g. coal use for primary energy, CO₂ emissions from the transport sector for CO₂ etc.), \bar{x} is the mean of a graph series over the model versions v and n is the number of graph series. This output index varies between 0 (no change) and 1.

Table 9 shows these input and output indices for each model version, and Table 10 compares them by dividing the input changes by the output index. Version 2.7 has very few changes from v2.1 (Table 9); so, the v2.7 results in Table 10 are not comparable with the other versions; it is a good example of how small parameter changes can substantially affect the results in a linear model. Had we instead divided

Fig. 12 CO₂ emissions from each sector in 2050 in each model version (bars). The lines show the CO₂ price with (blue) and without (red) elastic demands



outputs by inputs then the indices in such cases would be extremely high as the denominator would be close to zero; for comparison purposes, it is more useful that the index tends towards zero in such cases. In Table 10, the indices are presented for the three combinations of input statistics. The coefficient of variation is also calculated for each set of indices (excluding v2.7) to examine whether the similar numbers of changes to the inputs lead to similar output variations across all of the versions.

The indices in Table 10 are higher for v3.17 than other versions because the high number of input changes does not lead to relatively greater output variations. The primary energy and final energy indices are more consistent across versions than the CO₂ sectoral emissions, as shown by the lower coefficient of variations. The indices are most consistent between versions when all input changes (structural, system constraints and parameters) are included. This is an important finding that shows that all model changes, including parameter changes, can have an important impact on the model and should be included in the model archaeology statistics.

7 Discussion

We have defined a series of model archaeology metrics for both model inputs and outputs and applied them to the UK MARKAL model as a case study. In this section, we reflect on the value of model archaeology by considering what we have learned about UK MARKAL.

7.1 Balancing Model Detail and Complexity

The total number of diverse technologies in UK MARKAL has increased steadily over time but only by a total of 13 % over the 5 years, with 89 % of technologies in the most recent version also present in the first version. This is despite a technology turnover rate of 12 %/year, showing that many technologies have also been removed. The regular changes demonstrate a sustained commitment to continually improving the model, but there has also been a strong focus to avoid making the model over-complicated by adding too many technologies, thus avoiding the pitfalls described in Ref. [19]. Our metrics show that the technology diversity of the model has been reasonably well balanced in terms of sectoral emissions and energy demands since the first version. Perhaps of more concern is the steady rise in the number of model system constraints which could potentially overly constrain the model in the future.

The overall rate of RES structural change is not much lower than the rate of change of the technology parameter data, reflecting an experimental and incremental approach to model improvement. Changes have been very much focused on particular sectors (electricity, transport, residential and resources), which reflect the priorities of the UK government

Table 8 Impact of model input changes on outputs on each sector. Similar numbers in each ‘Indices’ column show that the number of input changes is well correlated to the magnitude of the output changes across several sectors

Sector	Inputs	Outputs		Indices	
		No CO ₂ constraint	With CO ₂ constraint	No CO ₂ constraint	With CO ₂ constraint
Electricity	28 %	0.24	0.78	1.17	0.36
Transport	14 %	0.60	0.75	0.24	0.19
Residential	30 %	0.28	0.83	1.07	0.36
Industry	1 %	0.28	0.27	0.05	0.05

as well as the interests of the modelling team (as measured by journal paper output). Changes have to some extent been driven by data availability; for example, electricity generation capital costs have been regularly updated, using reports that are frequently commissioned by the UK government, while residential heat changes have focused on energy efficiency updates and costs have remained unchanged.

Our ex post analysis of input statistics identifies several long-term cost and energy efficiency trends in different sectors. Overall, the UK MARKAL representation of the future energy system has become more pessimistic over time, with a trend towards higher technology costs and lower energy efficiencies particularly noticeable in the electricity sector. This is interesting for two reasons. First, it means that estimates of the impact of decarbonisation on the UK economy, which were carried out using version 2.7, would likely increase were the analysis repeated using parameter data from v3.26. However, the addition of elastic demands and new mitigation options since v2.7 enable the model to avoid some of these increased costs; so, it is not clear whether the overall decarbonisation cost has increased or decreased; Fig. 12 shows that the marginal CO₂ price is broadly similar for all of the versions. Second, it

Table 9 Changes in the model inputs and outputs in each version. “Structural changes only” includes only changes to the model RES structural changes. “+ system constraints” additionally includes system constraint changes and “+ parameter changes” averages changes across all three model archaeology metrics. See the text for an explanation of how these figures were derived

Version	Input changes			Output changes		
	Structural changes only	+ system constraints	+ parameter changes	Primary energy	Final energy	CO ₂
2.7	<1 %	<1 %	<1 %	0.51	0.06	0.12
3.17	41 %	53 %	42 %	0.38	0.17	0.32
3.22	12 %	16 %	15 %	0.23	0.08	0.36
3.25	4 %	20 %	18 %	0.26	0.13	0.37
3.26	4 %	4 %	14 %	0.15	0.06	0.07

raises the question whether similar increases would have occurred in other sectors if they had received the same attention as the electricity sector. It would be useful to understand whether the underlying causes of these changes have ramifications for other sectors.

7.2 Usefulness of Comparing Ex Post Inputs and Outputs

Comparing ex post inputs and outputs helps us to identify the sensitivity of each sector to changes and can therefore be used to more effectively target future model improvements. Although the primary and final energy consumptions do not greatly change between most versions, there are large variations within some individual sectors. For some sectors, for example electricity, there is a clear link between changes to input data and outputs. In contrast, for the bioenergy, transport and residential sectors, data changes in other sectors have an important impact on outputs. Changes to inputs do not always translate to changes in outputs; for example, the updates to the transport sector in v3.26 have little discernible impact on outputs; yet, the outputs from v3.22 are quite different to v3.17 despite almost no changes being made to v3.22.

The metrics that we have derived in Section 6 to quantitatively compare the impact of input changes on model outputs reflect these conclusions. While we find a link between the frequency of input changes within a sector and the magnitude of output variations in Section 6.1, there are still important output variations in sectors with few input changes because all of the sectors are affected to some extent by input changes in other sectors. This means that it is necessary that all sectors are accurately represented in the model in order to produce meaningful results. This characteristic of the energy system highlights the need for energy system models in addition to sectoral models but also limits the usefulness of sectoral input-output indices for model archaeology. We plan to extend these quantitative analyses in the future to better understand cross-sectoral linkages by identifying the sensitivity of output variations in each sector to input changes in other sectors.

Table 10 Indices linking input changes to output changes between model versions (calculated by dividing inputs by outputs). Results are presented for the three sets of input changes used in Table 9. The same outputs are used for all three input sets. Where numbers in a column have a similar

Changes in primary and final energy consumption between versions can be linked to changes in inputs but only if the minor changes in v2.7 are excluded. These metrics are most stable when all model input changes, including parameter changes, are incorporated into the input indices, showing that all model changes can have an important impact on the results and should be included in the model archaeology statistics. Yet, the metrics do not work as well for sectoral CO₂ emissions. Moreover, the higher indices for v3.17, which has a much greater number of input changes than the other versions, suggest that output variations might plateau once a threshold level of input changes is reached. It would be interesting to compare these UK MARKAL statistics with similar statistics from other energy system models to understand whether the linkages between inputs and outputs that we have found occur more generally across these types of model. We hope that other researchers will test the performance of the model archaeology metrics on their models so that refined metrics can be identified for widespread adoption.

Changes to the model paradigm can be just as important as changes to the input data; for example, the introduction of elastic demands to UK MARKAL reduced energy consumption in all sectors. It is important therefore to look beyond RES improvements to understand the limitations of and potential improvements to the model paradigm. For example, the UK MARKAL modellers have managed this challenge through the introduction of elastic demand, the creation of high-resolution spatial and temporal versions and through experiments using a hybrid macroeconomic version and a stochastic version.

Another benefit of the ex post analysis of inputs and outputs is a better understanding of the robustness of the model results. We can find out whether outputs are robust to changes to input data and assumptions (e.g. nuclear power) or whether outputs are simply consistent because that part of the model has not been updated in subsequent model versions. For example, industrial sector fuel consumption in UK MARKAL is similar for all model versions, but we do not

magnitude, this indicates that the number of input changes is correlated to the magnitude of the output variations across several versions. “cv” is the coefficient of variation across the versions except for v2.7

Version	Structural changes only			+ system constraints			+ parameter changes		
	Primary energy	Final energy	CO2	Primary energy	Final energy	CO2	Primary energy	Final energy	CO2
2.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
3.17	1.09	2.43	1.30	1.41	3.13	1.67	1.12	2.48	1.33
3.22	0.52	1.50	0.33	0.69	1.97	0.44	0.67	1.92	0.42
3.25	0.15	0.30	0.11	0.76	1.50	0.53	0.69	1.36	0.48
3.26	0.24	0.62	0.50	0.25	0.66	0.52	0.92	2.41	1.91
<i>cv</i>	<i>0.85</i>	<i>0.79</i>	<i>0.93</i>	<i>0.61</i>	<i>0.57</i>	<i>0.75</i>	<i>0.25</i>	<i>0.26</i>	<i>0.69</i>

know if this pattern is robust to alternative methods of modelling the sector because it is virtually unchanged in all versions.

7.3 Process of Model Evolution

The model archaeology metrics show that improvements to UK MARKAL have been implemented at a steady rate over the last few years. Yet, different sectors have been changed in different versions, and some sectors have not been changed at all, partly because the updates have been influenced by research and policy interests and by the availability of better data. Yet, despite these influences, there has been a clear strategy to avoid overcomplicating any of the sectors.

A number of research versions of UK MARKAL have been produced, but the changes have rarely been fed back directly into the production versions. This approach potentially reduces the transparency and repeatability of experiments using the research versions because the underlying production versions become obsolete. It also underlines how important it is to interpret model results within the context of a given project. Model archaeology can help bridge these difficulties by characterising the differences between production versions in terms of both inputs and outputs. Other researchers can then use this characterisation as a foundation to understand the differences between production versions and also the differences between a research version and the production version from which it was derived. The research versions tend to have much greater changes to the model structure than the production versions, and it would be interesting to apply model archaeology to compare the production and research versions of UK MARKAL in the future, to understand whether the results of research versions tend to diverge more as a result of these structural changes.

7.4 Broader Application of Model Archaeology Metrics

In this paper, we have described how model archaeology can be used to understand the evolution of an energy system model. Similar model archaeology metrics can also provide a qualitative and quantitative foundation for comparisons of different energy models more generally. Using modelling metrics on both an input and output basis, and critically taking into account the dynamic evolution of models, will significantly improve ex post model analyses.

Researchers could use the formal qualitative and quantitative model archaeology metrics to holistically analyse model inputs as well as to provide a fuller picture of the differences between models and the robustness of results to model improvements. In our experience, models that produce different results to the majority in inter-comparison projects can be less well regarded by the modelling community, albeit unfairly in some cases. The model archaeology metrics could potentially identify the causes of such differences and improve the usefulness of such comparisons.

Model archaeology can also help understanding the evolutionary process of separate production and research versions, such as for UK MARKAL, and can hence improve the transparency of research model studies. This directly links into calls for improved version control and quality assurance of models whose useful life is in years or decades [56].

Model archaeology compels the modeller to perform a systematic review of the model and the modelling process. From that perspective, useful outcomes will include areas of model fragility and scope for future improvements. It provides a theoretical foundation for characterising models, and we hope that other modellers can use these techniques to underpin future research proposals and projects.

Acknowledgments This study was supported by the UK Energy Research Centre (grant number NE/G007748/1), which is funded by the RCUK Energy Programme. The UK Energy Research Centre has supported the long-term development of the UK MARKAL model, without which many of the studies cited here would not have been possible. We acknowledge the efforts of the many researchers that have contributed to the development of UK MARKAL, particularly Steve Pye, Kannan Ramachandran, Gabriel Anandarajah, Will Usher and Paul Ekins. We thank two anonymous reviewers whose insights helped us to improve the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- DeCarolis, J. F., Hunter, K., & Sreepathi, S. (2012). The case for repeatable analysis with energy economy optimization models. *Energy Economics*, 34(6), 1845–1853. doi:10.1016/j.eneco.2012.07.004.
- Das, A., Rossetti di Valdalbero, D., & Viridis, M. R. (2007). ACROPOLIS: an example of international collaboration in the field of energy modelling to support greenhouse gases mitigation policies. *Energy Policy*, 35(2), 763–771. doi:10.1016/j.enpol.2006.03.004.
- Edenhofer, O., Knopf, B., Barker, T., Baumstark, L., Bellevrat, E., Chateau, B., et al. (2010). The economics of low stabilization: model comparison of mitigation strategies and costs. *The Energy Journal*, 31(Special Issue), 11–48. doi:10.5547/ISSN0195-6574-EJ-Vol31-NoSI-2.
- Pilavachi, P. A., Dalamaga, T., Rossetti di Valdalbero, D., & Guilmo, J.-F. (2008). Ex-post evaluation of European energy models. *Energy Policy*, 36(5), 1726–1735.
- Loulou, R., Goldstein, G., & Noble, K. (2004). Documentation for the MARKAL family of models. ETSAP, IEA. Available at: www.iea-etsap.org/web/Documentation.asp.
- Messner, S., & Strubegger, M. (1995). *User's guide for MESSAGE III*. Laxenburg: International Institute for Applied Systems Analysis. Available at: <http://webarchive.iiasa.ac.at/Admin/PUB/Documents/WP-95-069.pdf>.
- Howells, M., Rogner, H., Strachan, N., Heaps, C., Huntington, H. G., Kypreos, S., et al. (2011). OSeMOSYS: the open source energy modeling system: an introduction to its ethos, structure and development. *Energy Policy*, 39(10), 5850–5870. doi:10.1016/j.enpol.2011.06.033.

8. IEA. (2012). *Energy technology perspectives 2012: pathways to a clean energy system*. Paris: International Energy Agency.
9. DECC. (2011). *The carbon plan: Delivering our low carbon future*. London: Department of Energy and Climate Change.
10. Strachan, N. (2011). UK energy policy ambition and UK energy modelling-fit for purpose? *Energy Policy*, 39(3), 1037–1040. doi:10.1016/j.enpol.2011.01.015.
11. EIA. (2013). *AEO retrospective review: evaluation of 2012 and prior reference case projections*. Washington DC: U.S. Energy Information Administration.
12. Oppenheimer, M., O'Neill, B. C., & Webster, M. (2008). Negative learning. *Climatic Change*, 89(1–2), 155–172. doi:10.1007/s10584-008-9405-1.
13. Dodds, P. E., & Demoullin, S. (2013). Conversion of the UK gas system to transport hydrogen. *International Journal of Hydrogen Energy*, 38(18), 7189–7200. doi:10.1016/j.ijhydene.2013.03.070.
14. Jablonski, S., Strachan, N., Brand, C., & Bauen, A. (2010). The role of bioenergy in the UK's energy future formulation and modelling of long-term UK bioenergy scenarios. *Energy Policy*, 38(10), 5799–5816. doi:10.1016/j.enpol.2010.05.031.
15. Stirling, A. (2010). Keep it complex. *Nature*, 468(7327), 1029–1031. doi:10.1038/4681029a.
16. Huntington, H. G. (2011). The policy implications of energy-efficiency cost curves. *The Energy Journal*, 32, 7. doi:10.5547/ISSN0195-6574-EJ-Vol32-S11-2.
17. Förster, H., Schumacher, K., De Cian, E., Hübler, M., Keppo, I., Mima, S., et al. (2013). European energy efficiency and decarbonization strategies beyond 2030 — a sectoral multi-model decomposition. *Climate Change Economics*, 4(Suppl. 1), 1340004. doi:10.1142/S2010007813400046.
18. Varian, H. R. (1997). How to build an economic model in your spare time. In M. Szenberg (Ed.), *Passion and craft: Economists at work*. USA: University of Michigan Press.
19. Morgan, M. G., & Henrion, M. (1990). *Uncertainty. A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge: Cambridge University Press.
20. van Vuuren, D. P., Weyant, J., & de la Chesnaye, F. (2006). Multi-gas scenarios to stabilize radiative forcing. *Energy Economics*, 28(1), 102–120. doi:10.1016/j.eneco.2005.10.003.
21. Anandarajah, G., Strachan, N., Ekins, P., Kannan, R., & Hughes, N. (2009). Pathways to a low carbon economy: energy systems modelling. London, UK. Available at: www.ukerc.ac.uk/support/UKERC2050PathwaysLowCarbonEconomy.
22. Keppo, I., & Strubegger, M. (2010). Short term decisions for long term problems — the effect of foresight on model based energy systems analysis. *Energy*, 35(5), 2033–2042. doi:10.1016/j.energy.2010.01.019.
23. Lempert, R. J. (2002). A new decision sciences for complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3), 7309–7313. doi:10.1073/pnas.082081699.
24. DeCarolis, J. F. (2011). Using modeling to generate alternatives (MGA) to expand our thinking on energy futures. *Energy Economics*, 33(2), 145–152. doi:10.1016/j.eneco.2010.05.002.
25. Anandarajah, G., McDowall, W., & Ekins, P. (2013). Decarbonising road transport with hydrogen and electricity: long term global technology learning scenarios. *International Journal of Hydrogen Energy*, 38(8), 3419–3432. doi:10.1016/j.ijhydene.2012.12.110.
26. DTI. (2003). *Our energy future—creating a low carbon economy*. London: Department of Trade and Industry.
27. Anandarajah, G., & Strachan, N. (2010). Interactions and implications of renewable and climate change policy on UK energy scenarios. *Energy Policy*, 38(11), 6724–6735. doi:10.1016/j.enpol.2010.06.042.
28. Kannan, R., Strachan, N., Pye, S., Anandarajah, G., & Balta-Ozkan, N. (2007). UK MARKAL Model Documentation. Available at: www.ucl.ac.uk/energy-models/models/uk-markal/uk-markal-documentation.
29. Strachan, N., & Kannan, R. (2008). Hybrid modelling of long-term carbon reduction scenarios for the UK. *Energy Economics*, 30(6), 2947–2963. doi:10.1016/j.eneco.2008.04.009.
30. Usher, W., & Strachan, N. (2012). Critical mid-term uncertainties in long-term decarbonisation pathways. *Energy Policy*, 41, 433–444. doi:10.1016/j.enpol.2011.11.004.
31. Strachan, N., Balta-Ozkan, N., Joffe, D., McGeever, K., & Hughes, N. (2009). Soft-linking energy systems and GIS models to investigate spatial hydrogen infrastructure development in a low-carbon UK energy system. *International Journal of Hydrogen Energy*, 34(2), 642–657. doi:10.1016/j.ijhydene.2008.10.083.
32. McDowall, W., Anandarajah, G., Dodds, P. E., & Tomei, J. (2012). Implications of sustainability constraints on UK bioenergy development: assessing optimistic and precautionary approaches with UK MARKAL. *Energy Policy*, 47, 424–436. doi:10.1016/j.enpol.2012.05.015.
33. Usher, W., & Strachan, N. (2010). *UK MARKAL modelling—examining decarbonisation pathways in the 2020s on the way to meeting the 2050 emissions target*. London: UCL Energy Institute. Available at: <http://downloads.theccc.org.uk/s3.amazonaws.com/4th%20Budget/CCC%20MARKAL%20Final%20Report%20-%20UCL%20Nov10.pdf>.
34. Anandarajah, G., & McDowall, W. (2012). What are the costs of Scotland's climate and renewable policies? *Energy Policy*, 50, 773–783. doi:10.1016/j.enpol.2012.08.027.
35. Dodds, P. E., & McDowall, W. (2012). *A review of hydrogen delivery technologies for energy system models*. London: UCL. Available at: www.bartlett.ucl.ac.uk/energy/research/themes/energy-systems/hydrogen/WP7_Dodds_Delivery.pdf.
36. Dodds, P. E. (2013). *Review and disaggregation of the UK MARKAL residential sector heat provision to assess the prospects for using hydrogen in the built environment*. London: UCL Energy Institute. Available at: www.bartlett.ucl.ac.uk/energy/research/themes/energy-systems/hydrogen/h2fc-ucl-working-paper-1-dodds-residential-heat-techs-2013.
37. McDowall, W., & Dodds, P. E. (2012). A review of low-carbon vehicle and hydrogen end-use data for energy system models. London: UCL. Available at: www.bartlett.ucl.ac.uk/energy/research/themes/energy-systems/hydrogen/WP5_McDowall_EndUse.pdf.
38. Kannan, R., Ekins, P., & Strachan, N. (2008). The Structure and use of the UK MARKAL model. In L. Hunt & J. Evans (Eds.), *International handbook on the economics of energy* (pp. 285–310). Northampton: Edward Elgar Publishing.
39. Strachan, N., Pye, S., & Hughes, N. (2008). The role of international drivers on UK scenarios of a low-carbon society. *Climate Policy*, 8(Supplement 1), S125–S139.
40. HM Government. (2007). *Meeting the energy challenge*. London: Department of Trade and Industry.
41. DEFRA. (2007). *MARKAL macro analysis of long run costs of climate change mitigation targets*. London: Department for Environment, Food and Rural Affairs. Available at: www.defra.gov.uk/environment/climatechange/research/index.htm.
42. Strachan, N., Pye, S., & Kannan, R. (2009). The iterative contribution and relevance of modelling to UK energy policy. *Energy Policy*, 37(3), 850–860. doi:10.1016/j.enpol.2008.09.096.
43. Kannan, R. (2008). Role of electricity and hydrogen storage in low carbon energy system — Modelling in temporal MARKAL model. Presented at the *International Energy Workshop*, Paris, France, 30 June - 4 July.
44. Kannan, R., & Strachan, N. (2009). Modelling the UK residential energy sector under long-term decarbonisation scenarios: comparison between energy systems and sectoral modelling approaches. *Applied Energy*, 86(4), 416–428.

45. MacDonald, M. (2010). *UK electricity generation costs update*. UK: Brighton.
46. MacDonald, M. (2011). *Costs of low-carbon generation technologies*. UK: Bristol.
47. Strachan, N. (2011). Business-as-unusual: existing policies in energy model baselines. *Energy Economics*, 33(2), 153–160. doi:10.1016/j.eneco.2010.10.009.
48. HM Parliament. (2008). The climate change act 2008. In Department of Energy and Climate Change (Ed.). London, UK.
49. Hawkes, A., Narkeviciute, R., Morris, S., & Solano-Rodriguez, B. (2011). Pathways to 2050 – detailed analyses. London, UK: AEA Technology plc. Available at: www.decc.gov.uk/assets/decc/11/cutting-emissions/carbon-budgets/2270-pathways-to-2050-detailed-analyses.pdf.
50. Dodds, P. E., & McDowall, W. (2014). Methodologies for representing the road transport sector in energy system models. *International Journal of Hydrogen Energy*, 39(5), 2345–2358. doi:10.1016/j.ijhydene.2013.11.021.
51. Dodds, P. E., & McDowall, W. (2013). The future of the UK gas network. *Energy Policy*, 60, 305–316. doi:10.1016/j.enpol.2013.05.030.
52. Ekins, P., Keppo, I., Skea, J., Strachan, N., Usher, W., & Anandarajah, G. (2013). *The UK energy system in 2050: comparing low-carbon, resilient scenarios*. London: UK Energy Research Centre. Available at: http://www.ukerc.ac.uk/support/tiki-download_file.php?fileId=2976.
53. DECC. (2011). *Renewable heat incentive*. London: Department of Energy and Climate Change.
54. Moomaw, W., Burgherr, P., Heath, G., Lenzen, M., Nyboer, J., & Verbruggen, A. (2011). Annex II: Methodology. In O. Edenhofer, R. Pichs-Madruga, Y. Sokona, K. Seyboth, P. Matschoss, S. Kadner, et al. (Eds.), *IPCC special report on renewable energy sources and climate change mitigation*. Cambridge: Cambridge University Press.
55. DECC. (2012). *Digest of United Kingdom energy statistics 2012*. London: Department of Energy and Climate Change.
56. HM Treasury (2013). Review of quality assurance of Government analytical models: final report. London, UK. Available at: http://www.hm-treasury.gov.uk/d/review_of_qa_of_govt_analytical_models_final_report_040313.pdf.

For instructions about how to access the UK MARKAL model, please see the model website at: www.ucl.ac.uk/energy-models/models/uk-markal. Please contact the corresponding author for details about the versions of the model and the data used in this study.