A qualitative evaluation approach for energy system modelling frameworks

Wiese, Frauke; Hilpert, Simon; Kaldemeyer, Cord; Pleßmann, Guido

Published in:
Energy, Sustainability and Society

Link to article, DOI:
10.1186/s13705-018-0154-3

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Wiese, F., Hilpert, S., Kaldemeyer, C., & Pleßmann, G. (2018). A qualitative evaluation approach for energy system modelling frameworks. Energy, Sustainability and Society, 8(1). DOI: 10.1186/s13705-018-0154-3
A qualitative evaluation approach for energy system modelling frameworks

Frauke Wiese1*†, Simon Hilpert2,3†, Cord Kaldemeyer2,3,4† and Guido Pleßmann5

Abstract
Background: The research field of energy system analysis is faced with the challenge of increasingly complex systems and their sustainable transition. The challenges are not only on a technical level but also connected to societal aspects. Energy system modelling plays a decisive role in this field, and model properties define how useful it is in regard to the existing challenges. For energy system models, evaluation methods exist, but we argue that many decisions upon properties are rather made on the model generator or framework level. Thus, this paper presents a qualitative approach to evaluate frameworks in a transparent and structured way regarding their suitability to tackle energy system modelling challenges.

Methods: Current main challenges and framework properties that potentially contribute to tackle these challenges are derived from a literature review. The resulting contribution matrix and the described application procedure is then applied exemplarily in a case study in which the properties of the Open Energy Modelling Framework are checked for suitability to each challenge.

Results: We identified complexity (1), scientific standards (2), utilisation (3), interdisciplinary modelling (4), and uncertainty (5) as the main challenges. We suggest three major property categories of frameworks with regard to their capability to tackle the challenges: open-source philosophy (1), collaborative modelling (2), and structural properties (3). General findings of the detailed mapping of challenges and properties are that an open-source approach is a pre-condition for complying with scientific standards and that approaches to tackle the challenges complexity and uncertainty counteract each other. More research in the field of complexity reduction within energy system models is needed. Furthermore, while framework properties can support to address problems of result communication and interdisciplinary modelling, an important part can only be addressed by communication and organisational structures, thus, on a behavioural and social level.

Conclusions: We conclude that the relevance of energy system analysis tools needs to be reviewed critically. Their suitability for tackling the identified challenges deserves to be emphasised. The approach presented here is one contribution to improve current evaluation methods by adding this aspect.

Keywords: Energy system analysis, Model challenges, Open science, Open source, Energy modelling framework, Oemof
control and planning, dispatch and unit commitment, expansion planning, and energy market design, as well as environmental and social analysis of highly integrated energy systems. Energy system modelling software has been heavily discussed, and in recent years, model-based results have been criticised for the black box character of internal model logic and underlying assumptions [1, 2]. As a result, more researchers have opened their software and data [3] which improves transparency, enables reproducibility, and allows other people to re-use or build upon existing tools. Thus, a rough division into a group of closed (1st generation) and a group of open (2nd generation) energy system models and frameworks can be identified [4].

The diverse research questions associated with the transformation of energy systems can obviously not be addressed by one single model or approach. This is underpinned by the large amount of existing models and their differentiation along social, technologic, and economic lines.

In the following, we distinguish between the three terms model, model generator, and framework. Models are concrete representations of real-world systems (e.g. with a specific regional focus and temporal resolution). Such a representation may consist of multiple hard- or soft-linked sub-models to answer clearly defined research questions. Models can be built from model generators that allow to build models with a certain analytical and mathematical approach (e.g. by the use of a pre-defined set of equations and represented technologies). Finally, a framework can be understood as a structured toolbox including sub-frameworks and model generators as well as specific models (e.g. wind-feed-in models).

Although existing energy model and framework overviews are not comprehensive [3, 5], it is obvious that their number is growing. Multi-purpose model generators or frameworks such as MARKAL [6], TIMES [7], OSeMOSYS [8], PyPSA [9], or oemof [4] are important within the energy modelling community. In this context, it is crucial for users and developers to identify software that is fit for the intended purpose. Due to the nature of model generators and frameworks with their multi-purpose design and versatility, this task is not trivial. Hence, methods for quantitative as well as qualitative evaluation are important in terms of software selection. For this task, scientific model comparisons for specific models and model fact sheets as well as transparency checklists have been proposed (see section ‘Background and motivation’). However, there is a lack of comprehensive evaluation of energy modelling software with regard to their suitability for tackling described modelling challenges.

In this paper, we propose a qualitative evaluation approach as a step towards model generator and framework evaluation. To illustrate its application, we apply the approach to a 2nd generation energy modelling tool. Within a case study, the Open Energy Modelling Framework (oemof) is evaluated regarding its capability to address present and future challenges in energy system modelling.

First, we give a short overview on existing evaluation approaches in the ‘Background and motivation’ section. Then, we describe the ‘Method’ section to derive our evaluation approach. Subsequently, the ‘Challenges’ section is discussed and summarised in combination with the ‘Framework properties’ section. This forms the basis for the presented evaluation approach, for which the ‘Application procedure’ section is described. In the next step, the approach is applied exemplarily in the ‘Case study’ section. Finally, we discuss the proposed approach and general findings of the challenge-property-matching in the ‘Discussion’ section. The ‘Conclusions’ section summarises the main findings.

Background and motivation

An evaluation of energy system modelling software can be undertaken by quantitative, qualitative, or mixed methods. Quantitative approaches may be used to evaluate performance in terms of run-time or computational traceability as well as accuracy of results. The US Energy Modelling Forum has conducted model comparisons since the 1980s by looking at the foci of models, their internal logic and representation, and their results (see e.g. [10, 11]). One example is the ongoing project RegMex, comparing simulation pathways of renewable energy systems [12]. In this context, standard test cases serve as a common basis for model comparisons. A mixed quantitative and qualitative approach is used in [13], where the evolution of a model is characterised by comparing different model versions. In that paper, specific input- and output-related metrics are defined to allow for quantitative comparison.

For analysing aspects that cannot be expressed in numbers, qualitative methods can be applied. Systematic reviews of models and presentations of classification schemes [14–17] fall into this category and are important for modellers, model users, and decision makers to identify the potential application scope of a model. Similarly, qualitative model comparison helps to understand the details of and differences between models that are designed to answer similar research questions. Another qualitative approach is presented in [18]. In order to increase transparency of energy scenario-based studies, a transparency checklist is proposed. In addition to enhanced transparency, this list may also provide a basis for model comparison. Besides reviews and comparisons, a presentation of consistent argumentation provides possibilities for analysing modelling software with respect to...
for example practicability or the degree of openness. To
our knowledge, this kind of analysis has not been applied
to energy system modelling software. In particular, an
approach specifically designed for a qualitative assess-
ment of model generators or frameworks does not exist.

We know that on the one hand, literature that iden-
tifies challenges for energy system modelling exists and
on the other hand, model fact sheets characterise proper-
ties of models are available as well. However, a systematic
mapping of how framework properties can serve as solu-
tions to specific challenges is missing. An evaluation based
on the relationship between challenges and framework
properties could therefore facilitate progress in model
tool development with regard to the actual research
needs. Furthermore, an analysis focusing on model gener-
ators and frameworks is missing. The suggested approach
builds upon fact sheets and checklists as well as on chal-
lenge classifications, but combines both and lifts it from
model to framework level.

Method
First, a literature review is conducted to compile energy
system modelling challenges. Each of the five derived
challenges and respective underlying aspects are then
discussed and reasoned. Subsequently, framework charac-
teristics that have the potential to tackle the derived chal-
lenge classifications are summarised in a matrix which serves as a
template for the suggested evaluation. In the next step, the
application procedure is tested with a case study and sub-
sequently adapted. We illustrate the application procedure
of the suggested evaluation by including the case study in
this paper.

Challenges
Overview
Other authors have characterised the field of energy sys-
tem modelling and its models as opaque to outsiders
[13, 15]. One reason for this may be the broad def-
ition of the term energy system model. Depending
on the research question, energy system models may
range from detailed highly technical models of small
sub-systems to large macro-economic models covering
whole economies. Typical criteria for categorising models
are top-down (macro-economic relationship of compo-
nents) vs. bottom-up (technology specific) approaches, simulation vs. optimisation of the system, and partial
equilibrium (e.g. considering only the power sector) vs.
equilibrium (considering the whole economy) models
[20]. For a comprehensive description of the model (gen-
erator) landscape as well as of model topology, we refer to
existing reviews [14, 15, 21, 22]. We restrict our analysis
to general challenges and their respective aspects in the
field of energy system modelling. These challenges relate
to steps in the modelling process as described by [23],
ranging from the development of a mental model of an
energy system to the application of the model including
the communication of results.

Coming up with a classification scheme for energy
modelling challenges can be compared to proposing a
scheme for energy model classification with regard to
the generality of the categories. In case of energy system
models, various options for classification exist, though
there are ‘few models—if any—that fit into one distinct
category’ ([24], p.7). This is similarly true for categori-
sation of energy system modelling challenges. For our
analysis, we propose the five major challenge categories
complexity, uncertainty, interdisciplinary modelling, sci-
entific standards, and utilisation, which are characterised
by different relevant aspects, summarised in Table 1.

| Challenge       | Aspects                                      | Literature        |
|-----------------|----------------------------------------------|-------------------|
| Complexity      | Increasing sector coupling, high technical, temporal and regional resolution required, extensive input data pre-processing, extensive result data processing | [20–22]           |
| Uncertainty     | Epistemic, aleatory, linguistic, decision, planning | [23, 31, 39]      |
| Interdisciplinary modelling | Inclusion of the human dimension, energy-water-food nexus, common transdisciplinary understanding | [42, 44, 54]      |
| Scientific standards | Transparency, repeatability, reproducibility, scrutiny, scientific progress | [57, 58, 64, 65, 68] |
| Utilisation     | Usability, applicability, re-usability, result communication | [64, 72, 74, 77]  |

Generally, the relevance of a challenge for specific soft-
ware may vary as it is determined by the focus of the
underlying research question. The subsequent sections
provide a detailed description for each of the identified
challenges.

Complexity
The challenge category complexity with its main aspects
sector coupling, technical, temporal, or regional resolu-
tion, input data, and result processing is linked to the
challenges in the utilisation category. There exists a con-
tinuous trade-off between modelling complex interactions
with the required level of detail and keeping the model or framework simple and comprehensible for the recipients of the results and for the modellers themselves.

Diversification, distributed generation, and stronger integration of energy sectors with versatile interdependencies are growing challenges for energy system modellers. Considering the power-heat-transport nexus, integrated models nowadays play a decisive role in providing insight into different flexibility options, using excess electricity economically [26], and for meeting climate goals [27]. While a high spatial and temporal resolution is required to consider varying weather conditions and cover different flexibility options, spatial and temporal coverage is also necessary for analysing the long-term development of an increasingly interconnected power system. For instance, Despré et al. [20] conclude that long-term energy models would benefit from an improved representation of fluctuating renewable energy sources in the power sector. The growing requirement of flexibility particularly on the demand side (e.g. storage or demand side management) additionally increases modelling complexity in systems with high shares of renewables.

The increasing complexity of models is accompanied by a rising amount and complexity of input and result data. Data are crucial since their absence may hamper the development of new modelling techniques, as Krysiak and Weigt [28] argue in the case of demand side management modelling. Keirstead et al. [22] state that data availability is one challenge for (urban) energy system modelling. Acquiring or generating input data is not a trivial task, as it requires versatile software skills (e.g. geographic information systems, databases, reverse engineering) and may be linked to other sophisticated research areas (e.g. meteorology in the case of power production from wind turbines). Therefore, data processing to generate model input data is often not only one of the most time-consuming tasks of the whole modelling process but also adds to the complexity. Different kinds of input data from different sources need to be consistent, and their influence on the results have to be assessed adequately.

Regarding the output, models with high spatial, temporal, and technical resolution usually produce large amounts of result data that have to be analysed. Among other aspects, appropriate visualisation of multi-dimensional data (temporal, regional, unit-wise) is increasingly challenging with increasing dimensions of model complexity. Depending on the kind of application and question to be analysed, the processing of results may be a difficult and time-consuming task in itself. Even if the question to be analysed focuses on just one result parameter, the relation to other result parameters and the relation between varying input parameters and the result parameters of interest need to be checked thoroughly to grasp interdependencies.

Uncertainty

Uncertainty has already been identified as a challenge for energy system modelling decades ago [29]. Craig et al. [30] state that uncertainties in long-range energy forecasts are systematically underestimated. Uncertainty in terms of energy system modelling can be sub-classified into a number of aspects. Generally, literature has different scopes, approaches, and scientific backgrounds to classify uncertainty, resulting in different classification schemes [21, 23, 31]. Mirakyan and Guio miss a ‘common agreement on typology of uncertainty’ [23]. They propose a new framework that has a broader scope and a more detailed classification compared to uncertainties described by Pfenninger et al. [21] and Hunter et al. [31]. This framework for categorisation of uncertainty incorporates energy system modelling, decision making, and subsequent planning processes: (1) linguistic uncertainty, (2) knowledge or epistemic uncertainty, (3) variability or aleatory uncertainty, (4) decision uncertainty, (5) planning procedural uncertainty, and (6) level of uncertainty.

Even though not very often discussed in the context of energy modelling, the aspect of linguistic uncertainty (1) affects energy system planning and decision making based on model results. Linguistic uncertainty arises from natural language being vague and ambiguous, as meaning of words may change over time [32]. An illustrative example is the ambiguous use of the term model.

Knowledge or epistemic uncertainty (2) covers various levels of uncertainty related to context or framing, data, structure of a model, or framework, as well as technical and accumulated uncertainty that includes all other. Various examples for this type of uncertainty exist in literature, as this category covers a wide range. Assumed learning rates and consequently future costs (e.g. for renewable energy technologies) are decisive parameters for energy system models, as those often aim for minimal system costs. If not carefully chosen, biased results may lead to incorrect policy recommendations if they do not reflect the sensitivity of these assumptions [33]. Methods and key pitfalls of assessing future costs of energy technologies based on learning rates are an important topic among the research community [34, 35] that illustrates the importance of dealing with uncertainty related to assumptions and input data. Another problem related to uncertainty is associated with scenario development. Laugs and Moll show that most scenarios only represent a small bandwidth of possible pathways. This under-representation of extreme scenarios hampers the scientific discourses and ‘skews the overall outlook on possible energy futures’ [36].

Structural uncertainty has special importance for long-term planning models as these cannot be fully validated [37]. Although tackling structural uncertainty is tricky, one attempt is made by DeCarolis et al. [38], who explore
the near-optimal decision space with their technique *modelling to generate alternatives* (MGA) [39].

Variability or aleatory uncertainty (3) refers to ‘inherent variability manifested in natural and human systems’ [23]. It can also be referred to as random or stochastic. The aspect variability can be addressed with established mathematical methods. For example, the open-source model generator TEMOA applies stochastic programming [31] to deal with variability uncertainty. For deterministic models, other options, such as scenario and sensitivity analysis or Monte Carlo simulations are available. However, sensitivity analysis or stochastic programming counteract the challenge of complexity, as these measures are computationally expensive. Even if such approaches are applied under the reasonable assumption of increasing computational resources, missing regulatory certainty in combination with disruptive events can hardly be tackled by existing technical methods. Hence, policy makers need to be aware that reliable policies and regulatory schemes are crucial for the degree of reliability that advice derived from energy modelling can offer in the future. Instead of handling these uncertainties as practical constraints, they have to be analysed additionally (e.g. influence of temporal and regional resolution on results). This is important, as growing complexity of the modelled systems requires reducing model complexity. In turn, structural uncertainties of these simplified models increase. Connected to this issue are open questions that directly link to utilisation (e.g. ‘Is a model with unquantified structural uncertainties fit for a specific purpose?’).

Decision uncertainty (4) stems from decision makers with a different understanding and judgement of objectives and appropriate solutions and strategies [23]. For example, risk perception or the way of presenting model results to decision makers may affect their decision [32]. Availability of resources in terms of information and time to process it affect decision making as well [23]. According to Wardekker et al. [40], uncertainty perception varies depending on the way information is provided. This relates planning procedural uncertainty to the aforementioned aspect decision uncertainty.

Interdisciplinary modelling

The development of energy system models is typically undertaken from an engineering or economic perspective. Jefferson [41] argues that emphasising equations and economic theories prevents researchers from focusing on complicated factors and their future implications. Furthermore, Wiese [42] states that twenty-first century challenges need to include other perspectives than least-cost optimisation. As stated above, differences are inevitable between ideal results of optimisation models with one single rational decision maker and real-world developments with a multitude of heterogeneous stakeholders [43]. In addition to an increased complexity, this is also a challenge from an interdisciplinary point of view since modellers need to integrate perspectives that are not captured by standard economic or engineering approaches. However, if energy research is not undertaken in an interdisciplinary way, researchers ‘are not likely to grasp the problems, and thus the solutions to this challenging (energy) research space’([44], p.247).

It is common to utilise Integrated Assessment Models in climate change research [45]. Also for the energy field, Integrated Assessment Models like TIAM-WORLD exist [46].

Social and behavioural factors are important to assess the adoption of renewable technologies [47, 48] or the representation of consumers’ real behaviour in energy models [49]. For example, social acceptance has a relevant impact on grid and wind power expansion [50, 51]. Heinrichs et al. [52] combine a survey, a macro-economic input-output model, and an energy system model to assess phase-outs of coal-fired power plants in Germany. They conclude that integrated assessment of energy systems provides more robust results.

Attempts exist to capture the human dimension in energy system modelling by applying social science methods. But considering the strong interconnectedness of energy systems and society, social sciences are rather under-represented in contemporary energy research [53].

Another requirement in interdisciplinary modelling results from the strong interdependencies between the energy, water, and food sector. Granit et al. [54] argue that an increased understanding of the water-energy-food nexus is necessary to achieve sustainable development goals. They present first attempts for integrated tools and state that further cooperation between the modelling disciplines is required.

To comprehend the dimension of challenges in interdisciplinary energy modelling, one has to consider that finding a coherent terminology and taxonomy within one field is already complex. This is referred to as linguistic uncertainty in the ‘Uncertainty’ section. Between different disciplines, a lack of understanding due to different terms impedes a common understanding of energy systems.

Scientific standards

Complying with scientific standards includes the aspects transparency, repeatability, reproducibility, and scrutiny. These principles ensure that science moves forward and can perform course corrections through independent verification [55]. Beyond that, these are also fundamental for the societal progress, which depends on return of knowledge that has been publicly funded. Repeatability or the sometimes used term replicability describes the ability to repeat an experiment and come to the same results. In contrast, reproducibility means that results can
be repeated by a different researcher in a different computer environment [56]. Although definitions exist, these two terms are not always used with this clear distinction in literature.

Transparency of methods, code, and data lays the foundation for the other three aspects, as it is a precondition for building up on existing scientific work in the field of energy system modelling. However, Ince et al. [57] state that for computer science, transparency at all stages constitutes a basic condition for reproducibility. Even if this is fulfilled, reproducibility remains a challenging task due to hardware, software, and natural language related uncertainty. The common situation of constantly changing versions of energy system models and failure to describe these precisely when presenting results, adds another dimension to the reproducibility challenge [13]. As Pfenninger et al. [1] argue, full―meaning effective―transparency of energy system models is still hampered by different barriers. Specifically, the lack of open licences on the original sources of data is an obstacle for making model data publicly available. Moreover, a sparse or non-existing documentation of data makes it inconvenient for others to use these data.

To facilitate repeatable analysis, DeCarolis et al. [58] recommend five steps of best practices in energy economic optimisation model development. These steps, we argue, can and should be applied to every energy system model and to some extent to frameworks: (1) make source code publicly accessible, (2) make model data publicly accessible, (3) make transparency a design goal, (4) develop test systems for verification exercises, and (5) work towards interoperability among models. In fact, with today’s information technology, it has never been easier to comply with these recommendations. However, regarding data, significant barriers still exist as explained above. In context of code, progress can be observed. Source code of different model generators has been made publicly accessible in recent years (e.g. Balmorel [59], OSeMOSYS [8], TEMOA [31], calliope [60], PyPSA [9]). Meanwhile, up to now, 25 open energy models and frameworks are registered on the website of the open energy modelling initiative [3]. Contrary to increasing model transparency, publishing solely aggregated results of energy system models is still a common procedure. For instance, a list of models used in the UK shows that input data and code of the majority of models are not open [15]. As almost any result can be generated by modifying decisive input data, variables, or code, the common practice makes repeatable results impossible. Attempts exist to overcome these problems. Regarding data, the Dataverse project is one example of technical support in linking associated data with the published article [61].

While point three on the list (transparency as a design principle) has already been discussed above as the foundation, the fourth point (verification exercises) refers to the aspect of scrutiny. The importance of scrutiny for energy system modelling is (in this paper) mainly discussed not only in a technical sense but also on the societal level. Point five on DeCarolis best practice checklist addresses applicability and re-usability which is discussed related to utilisation.

On the technical level, scrutiny refers to identifying inconsistencies or faults (so-called bugs). Every computer model is prone to bugs, whereas the probability of these errors increases along with the complexity and size of the model. Detecting bugs is particularly vital in energy system modelling, as small errors may have great impact on the results. Johnson [62] highlights that peer-reviewed open-source software has significant advantages regarding bug detection. Besides this, Ndenga et al. [63] point out that the size of a community, i.e. users and developers, is one metric for bug reports.

On a social level, scrutiny refers to the detection of bias in model code and data. The possibility to scrutinise model results is essential for credibility [64], and the development of public trust in the modelling results, particularly as participation of society in the design of energy pathways, becomes increasingly important [65]. Methods for stakeholder participation in transition processes towards sustainability are available and applied [66] although simultaneously creating new challenges [67].

Being widely used for policy advice, the trade-off between being policy-relevant without being policy-prescriptive is vitally important for model-based research [68]. Though, Mai et al. [64, p.9] conclude that, accidentally or purposefully, all models incorporate biases. Going one step further, Biewald et al. [69] argue that value-laden and ethical issues cannot and should not be avoided in model-based studies, but assumptions based on ethical opinions should be communicated transparently, which can increase policy relevance of these studies. Similarly, Edenhofer and Kowarsch [70] state that value-neutral scientific recommendations for public policy are not possible. As model-based research has to deal with normative-ethical aspects, they suggests a new culture in academia that defines the role of modellers as cartographers of solution spaces. Detecting value-laden assumptions is even more difficult than detecting bugs, as software tests fail at this. Hence, again transparency of source code and data is pivotal for energy model use in policy advice and essential for complying with quality standards [71].

Although all discussed aspects refer to all computational intensive sciences, Pfenninger et al. [1] argue that energy policy research is still lagging behind other fields in terms of complying with scientific standards.
Utilisation

The aspects of the challenge framework and model utilisation are linked to growing model complexity. In the modelling process, three main groups of persons are involved: (1) developers, (2) users, and (3) decision makers. It is noteworthy that in some cases, these groups may not be completely distinct, as developers and users might be identical. Regarding the user/decision maker interface, the user needs to be able to explain the model logic and its effects on results to recipients of the results. The aspects usability and result communication are associated with the user/decision maker interface. The other two main aspects identified with regard to utilisation are applicability, that can be understood as a problem of ‘ease of use’ at the developer-user interface, and re-usability, that can be understood as ‘ease of adaptation’ at the developer/decision maker interface.

As models only produce useful information if the recipients understand the causal relations to a certain degree, there remains a trade-off between the level of complexity and the general usability. Bale argues that ‘[m]odellers need to engage with their beneficiaries from the outset so that models are properly scoped and fit for purpose’ ([72], p.157). Most notably, this is important as models are made for obtaining insight, not for generating numbers [73].

The struggle of finding a common language between developers, users, and recipients of their results has existed almost as long as the models themselves. In 1976, the Energy Modelling Forum was formed to ‘foster better communication between the builders and users of energy models in energy planning and policy making’ ([74], p. 449). Energy research is generally application-oriented, but stands out among other policy fields with vertical and horizontal complexity, entailed costs, and strong path dependency, energy models are indispensable for policy support [75]. However, the decision makers’ idea of useful information may differ significantly from those of the users ([64], p.9). This is a crucial point, as ‘[a] model is not fit for purpose if it is developed without sufficient critique of the motives for producing the model’ ([72], p.155). Therefore, the aspect communication of results is a crucial aspect of the modelling process. In particular, valuable information may not only be lost at the user/decision maker interface. To tackle this problem in operation research, the concept of model assessors, analysing, and evaluating models for decision makers has come-up a long time ago [76]. Additionally, Strachan et al. [77] propose further improvements, such as platform-based expert user groups for coordination and interdisciplinary external stakeholder review for energy system models.

Between developers and users, an easier and better understanding of framework or model mechanisms than at the developer/decision maker interface could be assumed. However, this seems not always to be the case. One example for differences in understanding models and results is the discussion about results from the NEMS model (see [78] and [79] for details). The usefulness of a framework increases if it can be applied to a diverse set of problems and by different researchers. Ideally, the expense of a developer for building up on an existing framework should be lower than the expense for building a new one from scratch.

In context of energy system modelling, it has been argued that ‘[s]ociety as a whole saves time and money if researchers avoid unnecessary duplication’ ([1], p.212). Considering the rising number of open energy models and frameworks for similar purposes [3], it yet seems that developers tend to rather develop a new framework than use existing ones. A reason for this may be the increasing complexity and different software skills required for adapting models or frameworks. Consequently, being open does not seem to be sufficient in terms of usability, even if a deep modelling understanding exists. Thus, the aspect applicability is also connected to scientific standards as it is vital for the repeatability and, more importantly, to the reproducibility of results.

The problem of how results are communicated is a recurring point in literature. Communication of energy system modelling results fails, when recipients only see concrete numbers (e.g. total energy system cost) as an outcome, though models should primarily be seen as a tool for understanding mechanisms and getting insights [70, 73]. Strachan et al. [77] proposed approaches to reinvent the modeller/policy interface for overcoming this problem. The communication with a recipient of model results cannot be tackled directly, but a framework can contribute to improving result communication by structured output that includes effects of parameters, ranges of uncertainties, and relative differences between scenarios, instead of results reduced to individual figures. Furthermore, extended use of pre-prints and discussions about results and methods within the community before the actual publishing process can be one step into the right direction.

Framework properties

We categorise framework properties that can contribute to tackling the challenges described above in three major categories: (1) open-source philosophy, (2) collaborative development, and (3) structural properties. More detailed characteristics of these three properties are listed in Table 2 and described below.

Free and open-source philosophy

Calls to ‘open up’ energy system models are getting louder, according to Morrison [80] motivated by the need for improved public transparency and scientific
reproducibility. Free software, open-source, and open-data are basic conditions for transparency and allow for repeatability, reproducibility, and scrutiny [58]. However, publishing undocumented source code of complex models still presents a serious obstacle to others. Therefore, code review, version control, and thorough documentation are important elements for effective transparency [81].

With a standardised input/output data format, simultaneous publishing of model source code and the corresponding documentation (including data and meta-data) is possible. Cross-platform data structure provides a flexible user interface and can contribute to lowering the entry barrier for new users. If supported by a clear version-control workflow, this allows for the release of monolithic model versions including data and documentation. In that way, scientific model results are transparent and enable reproducibility.

Policy measures and planning processes based on the results of energy system models cannot be affected directly by the modeller. However, an open-source and open-data approach enables decision makers and planners to obtain a deeper understanding of model results considering details of model inputs. This may enhance communication between modellers, decision makers, and other stakeholders.

### Collaborative development

Different important characteristics of frameworks originate from collaborative development. This kind of development is a new challenge within the field of energy system modelling. But if it is done, it can trigger a process of finding common definitions and a shared understanding of energy research-related problems.

With a collaborative concept, frameworks can contribute to a process of addressing linguistic uncertainty. Identifying common elements in energy system modelling can help to determine coherent terminologies. Here, experience from collaborative modelling is key for the necessary interface definitions of different existing models. Therefore, at least ambiguity is inherently tackled as developers have to agree on specific terms during the development process. A common terminology enables the different groups to communicate effectively.

In the process of collaborative development, multiple perspectives of developers with different backgrounds can decrease the risk of overlooking or omitting decisive features of energy systems. Developing a common understanding of interdisciplinary problems is not a trivial task, but a necessary basis for appropriate modelling. Here, collaborative development may play an important role in translating into interdisciplinary model development.

Additionally, a collaborative framework development and thus more people working with the code basis increases the probability of finding bugs [82]. This can also be integrated in a more structural way by standard test procedures before merging new developments into the master version. Test systems for verification exercises are one of the recommendations for repeatable analysis by DeCarolis et al. [58].

A collaborative framework development with developers from different backgrounds requires a high-quality documentation. This results in improved transparency for new developers and external users and thus supports scientific progress.

The experience developers collect in a collaborative development process, how to find common definitions etc., are a good foundation for collaborating in an even more interdisciplinary team. The resulting generic basis allows for an easy coupling of energy system model components with new model components of other research areas (e.g. components in water resource modelling, investment delays due to public acceptance, demand changes due to behaviour changes). This supports the interoperability of models, which is important for repeatable analysis [58]. The generic approach is part of the structural properties, which is explained in the following.

### Structural properties

Structural properties of frameworks decide how flexible energy system models can be created, adapted, and linked. If essential structural properties are shared, hard- and soft-linking of applications based on the same framework can be performed even with different modelling approaches or with models using different regional and temporal resolutions. As Trutnevye et al. [83] argue, this can be a key for energy system analysis.

A modular design—where each module has a certain degree of interdependence from the remaining part of the framework—increases applicability of a framework since new users can create applications based on the desired module without knowledge of the complete framework. A framework that is not restricted to a specific mathematical approach facilitates the integration of other modelling techniques. That could be for example agent-based models or methods to capture the human

### Table 2: Framework properties with respective characteristics that are decisive for tackling the challenges

| Property                      | Characteristics                                                                 |
|-------------------------------|-------------------------------------------------------------------------------|
| Free and open-source software  | Open-source, documentation, version control, openness of data, code review    |
| Collaborative development     | Consistency of terminology, developer perspective spectrum, interdisciplinar- |
| Structural properties         | Modularity of framework structure, object-oriented implementation, generic   |
|                               | concept of energy system representation, data model                          |

...
dimension which would thus support interdisciplinary modelling.

Overall, a generic basis in combination with a flexible programming language facilitates a modelling process for complex and changing systems. Generic classes facilitate the integration with other models.

An object-oriented approach generally provides a flexible interface for extensions. This supports the development of energy systems based on the same framework separately by different persons and to connect afterwards.

Platform-independent software increases the usability of a model. If a model is tested on the main operation systems (Windows, Linux, Mac OSX), the potential user target group is enlarged. Python is a common programming language for relatively new open frameworks and models [81]. It has the advantage that required Python versions, and packages can be installed in a specific environment on the machine of the user. This makes sure that the framework can be run with the working version independently from other Python installations on the user’s machine. Such a high re-usability and adaptability could save other resources (e.g. time) in terms of parallel work, especially when it comes to long-term projects with a great extent of interfaces between groups and work packages. This is in line with the argumentation of increased productivity through collaborative burden-sharing ([1], p.212).

Despite abovementioned problems of existing approaches to tackle variability uncertainty (see the ‘Uncertainty’ section), variability uncertainty can partly be addressed with incorporated tools for sensitivity analysis. Methods to explore a large space of parameter variations (i.e. scenario or sensitivity analysis) can be built on top of framework-based models. This is easier if a modular and generic structure allows for it. However, one has to keep in mind that uncertainty cannot only be fully tackled by an energy system framework with current methods but also depends on the regulatory consistency. Additionally, the trade-off between complexity and uncertainty has to be balanced by modellers and model users and does not fully depend on framework properties.

**Application procedure**

The result of our review of energy system modelling challenges on the one hand and framework properties influencing the capability of frameworks to tackle these on the other hand is summarised in a matrix (Fig. 1.)

This derived matrix can be used to evaluate energy system modelling frameworks or model generators regarding their capability to cope with present energy system modelling challenges.

The evaluation we suggest is made along the proposed challenge-property-matrix in the following steps:

- Quantify the characteristics of the framework’s properties in focus: no/not available (o) - partly (+) - strongly (++).
- Argue for each challenge: how does each characteristic partly/strongly address which aspect of the challenges?
- Quantify the contribution level for each characteristic-aspect pair: not addressed (o) - partly addressed (+) - strongly addressed (++). Each characteristic can contribute to tackling a challenge aspect with at most equal rating. For example, if characteristic documentation is only partly available, it can contribute at maximum partly to tackling challenges.

![Fig. 1 Challenges (with aspects)—properties (with characteristics) matrix](image-url)
• Check if the written argumentation supports the quantitative result.
• Optional: If the framework in focus has additional properties and/or characteristics relevant for the challenges, these might additionally be added to the matrix and evaluated with regard to their contribution in a second round.
• Summarise potential changes of the framework that would improve the contribution to the challenges.

In the following, the procedure is exemplarily applied to a case study.

Case study
First, oemof is shortly described with respect to the listed properties (Open Energy Modelling Framework). Then, as outlined in the section ‘Application procedure,’ oemof’s characteristics are checked for each challenge, and their contribution in tackling the challenge is debated (‘Evaluation’ section). Finally, the resulting matrix summarises the findings (‘Summary’ section).

Open Energy Modelling Framework
The framework itself and the characteristics we refer to in this section are described in existing publications [4, 84] and the online documentation of the framework [85]. In the following, additional literature is referenced where necessary. The framework has been developed for the analysis of energy supply systems considering power and heat as well as (prospectively) mobility. It consists of different libraries with defined interfaces for their combination. Applications depict concrete energy system models constructed from oemof libraries. Inside comprehensive models, specific parts of such an application can be developed flexibly by combining oemof libraries with external libraries. The core concept of oemof is based on a network structure which describes the general topology of an energy system.

Available applications built within oemof (e.g. renpassGIS [86], reegis [87], HESYSOPT [88]) demonstrate that the modular approach of the framework allows the creation of applications with very different objectives. The general description, the toolbox character, and the flexibility concerning temporal and spatial resolution makes oemof a framework instead of a model. It is implemented in Python using several packages (e.g. for data analysis, optimisation) and can optionally be combined with a PostgreSQL/PostGIS database.

As a first step of the evaluation process, the characteristics are quantified in Fig. 2.

Evaluation

Complexity
Due to its structural properties, oemof allows to create flexible energy system models which can be adapted and linked. For example, modelling strongly integrated energy systems is straightforward, due to oemof’s network structure. If, for instance, a specific sub-system should only appear in certain calculations, it can be connected and disconnected flexibly to a graph-based energy system representation with all its components depending on the requirements.

Additionally, generic classes can be used to easily integrate other models. This has, for instance, been tested with the PyPSA library [9]. Applications built in oemof have shown the integration of electricity, heat, and mobility as well as energy market simulation models [89] and power flow analyses [90].

The temporal and regional resolution are not fully addressed, as no specific methods are implemented within
the framework. Nevertheless, the problem of resolution is partly addressed. For example, the optimisation library \texttt{solph} provides time-step-flexible modelling with time-step-dependent weighting.

The object-oriented approach of \texttt{oemof} generally provides a flexible interface for extensions. Different applications based on \texttt{oemof} can be hard- or soft-linked even if using different regional or temporal resolution. Based on the underlying concept, incorporating new modelling methods is possible although not done yet (e.g. agent-based models based on \texttt{core} components).

Moreover, the framework provides a complete toolkit for modelling highly integrated, renewable-energy-based systems. Thus, not only optimisation models can be built but also input data such as feed-in or demand time series may be generated on the basis of \texttt{oemof} functions. Especially, the feed-in libraries allow for a high spatial and temporal resolution.

Overall, the underlying generic basis in combination with a flexible programming language facilitates the modelling process for complex and changing systems.

\textbf{Uncertainty}

With its collaborative concept and a group of developers with different backgrounds, \texttt{oemof} contributes to a process of addressing linguistic uncertainty. \textit{Epistemic uncertainty} related to model structure uncertainty is partly addressed by \texttt{oemof} as well due to the multiple perspectives of the developers. At the moment, the framework does not provide any functions tackling problems of variability uncertainty.

\textbf{Interdisciplinary modelling}

The provided framework does not directly address the aspect of taking down disciplinary walls between energy system modelling and other research disciplines. However, the concept allows to integrate other modelling techniques, i.e. approaches that suit interdisciplinary modelling.

\textbf{Scientific standards}

\texttt{oemof} is licenced under the GNU General Public License v3.0 and thus meets a basic standard in terms of transparency and allows for repeatability, reproducibility, and scrutiny. The developer group also aims at open-data, but that is not yet fully achieved.

Another element of effective transparency is the four levels of documentation: (1) comments inside the code explaining non-intuitive lines of code; (2) docstrings inside the source code describing how to use the various classes, methods, and functions; (3) higher level descriptions of possible interactions between different libraries or application-specific usage information; and (4) application examples especially useful to new users.

Transparency on application level is supported by a standardised input/output data format and functions for simultaneous publishing of model source code, data, and meta-data. The data structure is human-readable, spreadsheet-based, and thus cross-platform applicable, which can contribute to lowering the entry barrier for new users. The version-control workflow supports reproducibility of model results. Backward and forward compatibility is ensured as defined in the semantic versioning approach \cite{semver}.

Regarding test procedures, there is a set of continuously extended tests (e.g. results, comparison of lp-files for mathematical models), which must be passed before merging into the development branch and thus also into the master branch. In addition, the \texttt{oemof} development uses (in addition to the tests) pull requests, which require review approvals of at least one more developer.

\textbf{Utilisation}

Beyond general challenges of utilisation outlined in ‘\textit{Framework properties}’ above, \texttt{oemof} supports the energy system modelling community by providing a basis for model development that is highly reusable and adaptable. It can be applied on the main operation systems and is—with Python as the programming language—based on platform-independent software.

Also, the applicability of \texttt{oemof} models is improved by the underlying structure. Once this structure has been internalised by users and model developers, its usage and development is straightforward. The different layers of \texttt{oemof} are partly independent from each other, which enables new users to create applications without knowing all parts of the framework.

Moreover, the overall concept is consistent and the graph-based structure is in line with the code, data, and documentation. Thus, even complex cross-sectoral models or applications developed with another scientific scope can be understood quickly. Generally, a well-defined modelling workflow increases overall transparency. The problem of result communication is only addressed indirectly by structured output that enables graphs for relative scenario comparison. However, this could be improved by providing methods for stating uncertainty ranges and methods for example visualising how strong input parameter variations affect different output parameters.

\textbf{Summary}

Figure 3 summarises all challenges and properties with their specific contribution levels for the evaluated framework.

Important issues related to complexity are particularly addressed by \texttt{oemof}’s structural properties. Due to the generic code basis and the object oriented implementation, the modular framework allows modelling of
Fig. 3 | Visualisation of evaluation results. Derived from mapping challenges and their aspects with properties and their characteristics

integrated renewable energy systems, easy model linking as well as input and result processing.

Most aspects of uncertainty are not tackled by oemof, but the collaborative development and structural properties may reduce linguistic and structural uncertainty. Important aspects like variability uncertainty are not addressed. This may be improved in future versions.

As delineated in the ‘Evaluation’ section, oemof lays important foundations for interdisciplinary modelling, as the generic basis allows for modelling components that have their origin in other research areas. However, this has not been implemented in any oemof applications so far.

Being developed in an academic context, challenges related to scientific standards are addressed thoroughly with the free software and open-source philosophy. Collaborative framework development requires a high-quality documentation and improves transparency for new developers and external users. Moreover, potential bugs can be identified and fixed quickly, due to a growing community and direct feedback between users and developers. This level of addressing is underlined by being in compliance with the best practice recommendations of DeCarolis et al. [58].

Regarding challenges in terms of utilisation, oemof’s philosophy constitutes an important precondition for tackling these. Effective transparency at all stages is crucial for communication of results, as well as for application building and re-usability of models. Similar to uncertainty, we find that frameworks could support tackling the aspects of the challenge utilisation to some extent. However, for example result communication, changes in debate culture would be required to fully make use of this provided support by energy modelling frameworks.

From the evaluation, we conclude that challenges related to complexity and scientific standards are strongly tackled. In contrast, uncertainty is not addressed at present, as major aspects of this challenge are not sufficiently considered. Regarding the challenges utilisation and interdisciplinary modelling, we argue that oemof captures these partially.

Discussion
This described evaluation approach provides a structured, hands-on procedure that comes with different disadvantages and advantages.

Due to the intensive analysis of framework properties required in the evaluation process, the addressees of this approach are rather developers or experienced users of the respective frameworks. This has been affirmed by the case study, which required specific experience which is difficult to derive from documentation only.

However, the general scheme can contribute to broadening the scope of a developer team applying it. The method of matching properties of a framework with challenges that need to be covered forces developers to think outside of the box of their own framework. Energy software development work is thus augmented by a societal perspective. Thus, it can be an assisting tool to relate one’s own work to the energy system analysis field. Furthermore, it can assist in identifying potential improvements to increase the relevance of the framework. We argue that the evaluation approach may also be applied for framework comparisons, given that people with in-depth knowledge of the respective frameworks are involved.

Advantages of the proposed approach lie in its simplicity and flexibility. Modellers can reflect modelling tools.
in a structured way regarding the selected challenges and properties. For a framework comparison with a special focus, additional challenges or properties could be added to the matrix.

However, the step from the qualitative description of the properties to their evaluation, as well as the matching of the contribution level of the characteristics to the challenges could incorporate subjective bias and vagueness. The potential bias could be lowered by a review of the valuation step from argumentation to the filled matrix by modellers new to the framework in focus. In the presented version of the evaluation approach, no weighting exists. Thus, results cannot reflect the importance of an individual aspect or characteristic. This would be a potential improvement.

Beyond the specific case study and the evaluation approach itself, the detailed discussion of challenges and properties in this paper identifies general issues in energy system modelling. Some approaches for tackling challenges may counteract each other, such as complexity and uncertainty. To allow for rather sophisticated methods like stochastic programming, more research in the field of complexity reduction within energy system models is needed.

Furthermore, some challenges cannot be tackled solely by the framework approaches. Among these are problems of result communication and interdisciplinary modelling. We argue that these must be primarily addressed through changes in communication and organisational structures, thus, on a behavioural and social level. Additionally, we suggest to increase efforts in evaluating if a model is fit for purpose and what its results may reveal—or may not reveal—for each model application. Here, collaborative and interdisciplinary modelling can be a valuable method supporting this process.

Conclusions
Although different approaches for evaluating and comparing models have been proposed, systematic evaluations of frameworks are under-represented. This paper presents an evaluation method for energy system modelling frameworks regarding their capability to address present challenges in energy system modelling. The application of our approach in a case study demonstrates its general capability to assess energy system modelling software in terms of present and future challenges. Advantages of the evaluation approach lie in its simplicity, flexibility, and transferability, whereas disadvantages are mostly due to potential subjectivity and bias.

In addition, the detailed mapping of challenges and framework properties reveals that an open-source approach is a fundamental condition for complying with scientific standards. Openness of a framework is also advantageous for its utilisation, meaning (re-)usability, and applicability. However, whether or not advantage can be taken from improved energy system modelling software is partly decided on social level. This is especially true for result communication and interdisciplinarity.

Acknowledgements
We would like to thank the working groups of the Center for Sustainable Energy Systems in Flensburg and the Reiner Lemoine Institute Berlin. Special thanks go to Birgit Schachler for valuable feedback on the manuscript, to Miriam Leich for proofreading, and to the oemof developing group for engagement in developing oemof and discussions regarding the evaluated framework.

Authors’ contributions
SH and CK performed the qualitative analysis of the framework (case study) in close cooperation and intensive discussion and contributed to the design of the paper; FW originally initiated the paper and contributed to the design of the study, to the review of energy system modelling challenges, and to the properties and the methodology. GP contributed to the review of challenges, in particular the challenge of uncertainty; all authors engaged in the initial design of the analysis; FW conducted the restructuring of the paper and the revision of the evaluation approach. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1 Department of Management Engineering, Technical University of Denmark, Productionstovet, Building 426, 2800 Kgs. Lyngby, Denmark. 2 Center for Sustainable Energy Systems (ZNES), Flensburg, Germany. 3 Department of Energy and Environmental Management, Europa-Universität Flensburg, Auf dem Campus 1, 24943 Flensburg, Germany. 4 Department of Energy and Biotechnology, University of Applied Sciences Flensburg, Kanzleistraße 91-93, 24943 Flensburg, Germany. 5 Reiner Lemoine Institut gGmbH, Rudower Chaussee 12, 12489 Berlin, Germany.

Received: 31 October 2017 Accepted: 13 March 2018
Published online: 16 April 2018

References
1. Pfenninger S, DeCarolis J, Hirth L, Quolin S, Staffell I (2017) The importance of open data and software: is energy research lagging behind? Energy Policy 101:211–215. https://doi.org/10.1016/j.enpol.2016.11.046
2. Pfenninger S (2017) Energy scientists must show their workings. Nature 542:393. https://doi.org/10.1038/542393a
3. Openmod Initiative (open energy modelling initiative) open models. http://wiki.openmod-initiative.org/wiki/Open_Models. Accessed 24 Jun 2016
4. Hilpert S, Kaldemeyer C, Krien U, Günther S, Wingenbach C, Plessmann G (2004) The open energy modelling framework—an approach to facilitate open science in energy system modelling. Submitted to Energy Strategy Reviews
5. RLTOKNI KTH Energy Modelling Platform Europe. http://www.energymodellingplatform.eu/resources.html. Accessed 18 Dec 2017
6. Loulou R, Goldstein G, Noble K Documentation for the markal family of models. Technical report. Energy Technology Systems Analysis Programme (ETSAP)
7. Loulou R, Remme U, Kanudia A, Lehtila A, Goldstein G (2005) Documentation for the times model - part I –78. Technical report. Energy Technology Systems Analysis Programme (ETSAP)
8. Howells M, Rogner H, Strachan N, Heaps C, Huntington H, Kyriakos S, Hughes A, Silveira S, DeCarolis J, Bazilian M, Roehl A (2011) Omerosys: the open source energy modeling system: an introduction to its ethos,
90. Müller UP, Cussmann I, Wingenbach C, Wendiggensen J (2017) AC Power Flow Simulations within an Open Data Model of a High Voltage Grid (Wohlgemuth V, Fuchs-Kittowski F, Wittmann J, eds.). Springer, Cham. https://doi.org/10.1007/978-3-319-44711-7_15
91. Preston-Werner T Semantic Versioning 2.0.0. https://semver.org/. Accessed 12 Jan 2018