Review

Complexity profiles: A large-scale review of energy system models in terms of complexity

Elias Ridha, Lars Nolting *, Aaron Praktiknjo **

RWTH Aachen University, Institute for Future Energy Consumer Needs and Behavior (FCN), Mathieustr. 10, 52074, Aachen, Germany

ABSTRACT

Energy systems are becoming increasingly complex as developments such as sector coupling and decentral electricity generation increase their interconnectedness. At the same time, energy system models that are implemented to depict and predict energy systems are limited in their complexity due to computational constraints. Thus, a trade-off has to be made between high degrees of detail and model runtimes. As a first step towards efficiently managing the complexity of energy system models, we examine the relationship between the purpose of models and their complexity. Using fact sheets on 145 models, we manually cluster these models based on their purpose and underlying research questions. Further, we conduct mathematical clustering using several clustering methods to investigate the reproducibility of our results. For our study, we define the complexity of a model as the level of detail in which it represents reality. We distinguish the level of detail into the four dimensions of temporal, spatial, mathematical and modeling content complexity. The differences between the clusters found in these dimensions are verified statistically using confidence intervals. 112 out of 145 models can be allocated to one out of four major clusters possessing clearly distinguishable complexity profiles: unit commitment, electrical grids, policy assessment, and future energy systems. In each of these profiles, high complexity in one dimension or subdimension is compensated by low complexities in other dimensions. We therefore conclude that when creating a model, modelers allocate complexity in order of priority on those features and properties that are particularly important for fulfilling the model's purpose. Our results provide a necessary basis for the emerging field of complexity management in energy system modeling and are therefore of high interest for the scientific community and the interpreters of model results such as decision makers from policy and industry.

1. Introduction

Computer models play a decisive role in understanding energy systems. They serve a variety of purposes, such as understanding load flows, determining the effects of energy policy measures and comparing chances and risks of different technologies. This broad variety of purposes has led to the development of an equally broad variety of energy system models.

Traditionally, energy systems have often been analyzed proprietarily [1] and sharing methods and best practices has only recently become a focus of the modeling community (see e.g. the IEA-ETSAP Community [2]). In our experience, researchers tend to build their own models for personal or small-scale use and subsequently employ them to answer as many research questions as possible. While modelers generally are interested in having models that are as accurate as needed while being as simple as possible, to our knowledge, systematic complexity management is currently not at the forefront of modelers' efforts. However, the increasing complexity of energy systems [3] causes needs for the holistic complexity management of energy system models.

Therefore, our work aims at shifting the focus from the research questions that can be answered using a given model to the questions: "*Which qualities does a model need in order to answer a particular research question? In which dimensions does it need to be complex and in which dimensions is a lower degree of detail sufficient?*" In order to do this, we analyzed the status quo of complexity management in energy system modeling, examining whether models used for similar purposes share similar qualities regarding complexity (so-called complexity profiles). To carve-out current approaches of energy system modelers towards complexity management, we define different dimensions of complexity

---

* Corresponding author.
** Corresponding author.
 E-mail addresses: lnolting@eonerc.rwth-aachen.de (L. Nolting), apraktiknjo@eonerc.rwth-aachen.de (A. Praktiknjo).

Abbreviations

GUI	Graphical user interface
MCA	Multiple correspondence analysis
MODEX	Model experiments
SSD	Sum of squared distances

---

and cluster existing energy system models according to these dimensions. To use the most relevant data, our clustering method focuses on the predominant complexity qualities and approaches of the models.

Our work is structured as follows: section 2 is dedicated to the relevant literature background, the subsequent section 3 serves to demonstrate the data used for our analysis, in section 4 we introduce our methods, the results are then shown in section 5, we provide a discussion in section 6 and conclude this paper in section 7.

2. Literature background

As the model landscape is becoming more diverse, model comparisons and overviews have received considerable attention. The first comprehensive model classification scheme was developed by van Beeck in 1999 [4]. The scheme lists several ways of classifying models, such as the model purpose, the modeling assumptions, its approach and methodology and its sectoral, temporal and geographical coverage. Many later attempts at classifying models rely on criteria similar to those by van Beeck. However, most models are difficult to allocate to strict classes [5].

Most recent attempts towards classifying models focus on specific subspaces of the modeling landscape [6], falling broadly into one of two groups: they focus either on models with a specific purpose (often, this is energy policy analysis or the analysis of national energy systems) or on models incorporating specific technologies (e.g. electric vehicles). In addition, there are also more general overviews over the model landscape.

Among comparisons of models with a specific purpose are the works of Möst and Keles [7], Savvidis et al. [8], Lopion et al. [12], Weijermars et al. [9], Pfenninger et al. [10], Fisher et al. [11], Ventosa et al. [12], and DeCarolis et al. [13]. The latter focus on "economy optimization models" – models used to generate insights about energy economics on regional and bigger scales. They evaluate twelve models with regard to their openness and reproducibility of modeling results. Möst and Keles [7] gather data on eight stochastic models for electricity market prices. Savvidis et al. [8] examine 40 models in order to find models able to answer questions relevant to energy policy. Lopion et al. [12] investigate models of national energy systems covering all energy sectors. They give a history of energy system modeling and highlight current trends and future challenges for energy system models. The purpose of their comparison is to help analysts in their choice of model and the scheme developed resembles the one by van Beeck [4]. The criteria used include the models' temporal and spatial horizon as well as their methodology and modeling approach. Weijermars et al. [9] delineate modeling approaches suitable for determining future energy mixes. They distinguish between six major approaches, such as energy consumption extrapolation and scenario analysis, and name relevant models in these categories. Pfenninger et al. [10] aim to give a guideline as to which models are particularly suited for the analysis of the challenges recent energy market and energy technology developments pose. Hence, they do not develop a model taxonomy, but discuss the current paradigms and challenges of energy modeling and how different approaches deal with these. One of the challenges they identify is complexity. While they explain how some modeling approaches deal with the complexity inherent in energy systems, the purpose of their work is no to analyze different energy modeling tools with regard to their complexity. Fisher

et al. [11] compile an overview of modeling tools suited for modeling US states' energy systems in compliance with the US Clean Power Plan of the Obama administration. Their goal is to give those tasked with creating transformation plans for US states' energy systems an overview over the tools at their disposal. Ventosa et al. [12] develop a model taxonomy for electricity market models. They categorize models according to their mathematical approach (i.e. optimization, simulation and equilibrium models), including detailed sub-categories, and present examples of models developed in these categories. In addition, they find further distinguishing attributes (e.g. the degree of competition modelled and the models' time scope) and discuss the three approaches with regard to these attributes.

Other model comparisons focus mainly on models dealing with specific technologies and their effects. Foley et al. [14] analyze six electricity system models in detail in order to help modelers with their choice of model, while Timmerman et al. [15] examine which energy system models can be adapted for modeling decarbonized industrial parks. Mahmud and Town [16] focus on the transport sector, comparing models implemented to investigate the effects of electric vehicles on power distribution networks. They identify 125 modeling tools, 44 of which they study in detail. Ringkjøb et al. [6] examine 75 models suitable for analyzing energy systems with high shares of renewables, using three main criteria: the models' "general logic" (i.e. top-down vs. bottom-up modeling, the models' purpose and their methodology), their spatial and temporal resolution and the technological and economic aspects included in the models. This scheme of categorization also resembles the one developed by van Beeck [4]. Connolly et al. [17] pursue a similar goal, focusing on models used for analyzing the integration of renewables into energy systems. They create a questionnaire directed at modelers, covering a broad range of attributes, including the models' users, their applications and some model properties. Their analysis covers 37 models.

Lastly, there exists a number of analyses that are part of neither of the two groups named above. One of these is Hall and Buckley [18]. They aimed at giving an overview over the general model landscape in the United Kingdom and helping analysts choose a model. Hence, their comparison is restricted geographically in terms of where a model is used rather than thematically. The classification scheme that is applied to 22 models again is similar to the one used by van Beeck [4] (e.g. some of its categories are modeling approach, methodology, and technological detail). Another general overview is offered by Jebaraj and Inian [19], who distinguish six types of energy system models by their approach and provide a chronology of models developed in these categories. Després et al. [20] aim at reconciling the differences between broad energy system models and detailed electricity system models in order to make use of both approaches' strengths. In order to do so, they propose a methodology to describe both model types. Their typology includes the models' "general context and positioning" [20] (e.g., their mathematical approach and the modelled energy systems), their spatial and temporal resolution and their technical and economic features (e.g. whether electricity transmission is considered).

The model comparisons, typologies and overviews named above have been limited in scope and in sample size. Their purpose was often to give an overview of existing models in order to help analysts choose a model. The aim of our study, however, goes beyond providing an overview over existing models and their suitability for addressing specific research questions. Rather, we focused on the models' complexity as introduced below. In the grand scheme, it is our intention to find out how complex a model has to be depending on its purpose. As a first step towards this understanding, we examined in which ways models with different purposes differ in their complexity. We believe that our novel approach contributes to existing literature and might improve energy system modeling in future.

2.1. Defining complexity

In order to examine the complexity of energy system models, it is necessary to define the term *complexity*. Different quantitative definitions of complexity exist in a variety of fields. They are based on discipline-specific understandings of complexity. Additionally, there is an interdisciplinary, broad and qualitative notion of complexity that stems from the study of complex systems. Understanding and describing the qualities and behaviors of complex systems is the task of complexity research, an interdisciplinary field, the findings of which are applicable to energy systems [3]. Complex systems possess a number of properties that distinguish them from other systems [3,21]:

- *Agents*: Agents are the actors in a system, making decisions based on their own motivations. Those need not necessarily be identical to the system's motivations and goals. Agents can learn and interact with other agents.
- *Networks*: Physical and non-physical networks connect agents. The connections between agents can vary in direction and strength.
- *Self-organization*: The system creates a structure autonomously based on its agents' behavior. As a result, the system as a whole develops into a certain direction without needing a singular authoritative agent guiding it.
- *Path-dependency*: A system's state is always partially a result of its past [22]. For example, lock-in effects can appear, rendering it difficult or impossible to return to an earlier state [23].
- *Emergence*: The system as a whole possesses properties and displays behavior that cannot be explained based on the properties and behaviors of its elements, solely.
- *Co-evolution*: Complex systems interact with other systems. Several forms of interactions (e.g. competition, interdependency) can be present at the same time.
- *Adaptability*: The system as a whole keeps its identity despite its elements changing.
- *Non-linearity*: A complex system's behavior is highly dependent on its environment. Even small changes in input can lead to drastic changes in output [22,24].

Specific fields' understandings of complexity and complex systems draw upon the findings of complexity research, often highlighting one or several of the properties named above (e.g. Ref. [25–28]). Thus, the concept of complexity can also be transferred to the field of energy system analysis. For example, households, energy producers, TSOs and governmental institutions constitute agents with differing and sometimes contradictory motivations. Lock-in effects take place in energy systems as well: the current global energy system, relying on fossil fuels to a great extent, displays inertia. This complicates the transition to an energy system with significantly reduced green-house gas emissions, an effect that has been named carbon lock-in Ref. [29]. Further examples of complex properties in energy systems including a detailed explanation are given by Bale et al. [3].

2.2. Complexity in energy system modeling

A model represents a real system in a simplified way as to allow understanding the system. By definition, a model has a purpose. This purpose in turn determines which parts of the system are modelled to what extent [30]. The areas in which a model has to be particularly detailed and thus complex depend on the research questions the model is supposed to answer. That is why we define the complexity of an energy system model as the level of detail with which it represents the real system.

We distinguish between four dimensions of complexity which we base on the works of Senkpiel and Winkelmüller [5,31]. Table 1 shows the four categories and their main properties.

Table 1
Complexity dimensions.

Complexity dimension	Properties
Mathematical	Mathematical model class, approach to uncertainty
Temporal	Temporal resolution and horizon
Spatial	Spatial resolution and horizon
System scope	Energy supply sectors, demand sectors, primary energy carriers, technologies for generation and conversion, technologies for energy transportation, storage technologies, dynamics[a]

[a] Dynamics refer to temporally interdependent developments such as learning curves and cost degression that can be included in or excluded from the modelled system scope.

- *Mathematical complexity*: The mathematical model class of the modeling approach used (such as linear, mixed-integer linear, non-linear, etc.) and its ability to reflect stochastic behavior of the represented system.
- *Temporal complexity*: The model's temporal resolution and horizon.
- *Spatial complexity*: The model's geographical resolution and horizon.
- *System scope*: The system's parts that are modelled and the depth of their representation.

Recent approaches towards a better management of complexity in energy system models can be found in Priesmann et al. [32], focusing on different complexity settings of energy system optimization models and Nolting et al. [33], focusing on metamodeling approaches.

3. Data

The data used for our investigation stems from the MODEX (Model Experiments) project. *Projektträger Jülich*, a German institution concerned with funding public research, invited modelers to respond to a survey regarding their models' properties. The data was collected in so-called fact sheets summarizing the main attributes of the listed models.

3.1. The MODEX fact sheets

The MODEX fact sheets contain data on approximately 150 models, the vast majority of which are contributions from Germany. There were 149 attributes examined. As the survey was posed in a multiple-choice format, a question typically correlates to several attributes. For example, there are five attributes regarding the models' temporal resolution ("annual", "hour", "15 min", "1 min", "other"). Some of the multiple-choice options were accompanied by a commentary field intended for detailed explanations. A more detailed overview on the attributes asked and the structure of the fact sheets can be found in Appendix A.

As shown in Table 2, the majority of the attributes (98) relate to the models' complexity. These attributes can be sorted into four categories corresponding to the four dimensions of complexity explained above. Four attributes related to the models' purpose. The remaining 47

Table 2
Number of attributes in the MODEX model list.

Category	Number of attributes
Complexity	
Energy system modelled	59
Mathematical properties	17
Spatial properties	13
Temporal properties	9
Thematic focus	4
Other	
Licensing	10
Programming	17
General information	20

attributes concerned the models' licensing and programming, the modelers themselves, their institutions and other general information.

3.2. Models contained in the MODEX fact sheets

Fig. 1 illustrates the number of developers involved in creating the models, number of users, and the existence of a graphical user interface (GUI), respectively. More than 80% of the models are created by a small team consisting of no more than ten persons and more than 60% comprise of a maximum of ten users. Only 3% of the models are used by more than 100 users. This indicates that the majority of the models represented in the MODEX fact sheets are small ones, likely developed for use in a single research facility. This is in line with the share of models that possess GUI. Only 40% of the models employ a GUI for all parts of the model or will be updated in such a manner. The rest of the models do not possess a GUI, requiring not only programming skills but also in-depth knowledge of the models themselves in order to be useable. This supports our initial assessment that the modeling landscape predominantly consists of proprietary models intended for use in the research facility where they are developed.

4. Methods: model clustering

In order to examine the connection between the research question to be answered using a specific model and the necessary complexity level, we used several clustering techniques for the models listed in the MODEX fact sheets. After conducting initial data preparations and necessary data cleansing, we manually carved-out groups of models that serve similar purposes (i.e. manual clustering). We then examined the complexity properties of the models and created a complexity profile for each of the clusters. Finally, we re-clustered the models algorithmically, using several algorithms in order to verify the results of the manual clustering process. The different steps of the process chain are described in further detail in the following.

4.1. Manual clustering

During the manual clustering process, the MODEX list's attributes "*methodical focus*", "*primary purpose*", "*primary outputs*" and "*example research questions*" were consulted. We removed models that did not contain information in these fields from the dataset. Each of the remaining 145 models' primary foci was inferred from the attributes named above. Then, we grouped the models according to their primary focus, leading to 10 clusters of different foci. Afterwards, we re-examined the models' thematic foci iteratively, assessing whether another cluster provided a better fit.

4.2. Complexity properties and profiles

The four biggest clusters contained a majority of the models (112 out of 145 models). These four clusters' complexity properties were then investigated and compared to those of all 145 models. For each attribute, we calculated the percentage of models in a cluster that support or possess it (e.g. "*xx% of the models from Cluster A* support *an hourly temporal resolution*"). In order to do this, we translated information from the MODEX fact sheets to binary values, i.e. reducing detailed textual explanations to either "*yes (= 1)*" or "*no (= 0)*". In doing so, multiple "yes (=1)" answers for a model in one category were counted several times. E.g. if a model can be run as a LP and a MILP, we processed both anwers. We used the percentages calculated to compare the clusters with each other and with the sum of all models examined. Since the models in the MODEX list constitute a sample that might misrepresent the general population of models, we verified these values statistically by calculating confidence intervals on a 95% significance level. There are several ways to generate confidence intervals for binary data. Following the method that Galvin [34] suggests, we chose *Wilson Score intervals*.

Finally, the clusters' complexity properties were aggregated into complexity profiles. The profiles detail the areas which a cluster's models are particularly complex in compared to other clusters.

4.3. Algorithmic clustering

The manual clustering process performed is based upon a subjective assessment of a model's focus. In order to further verify and assess the quality of the models' cluster allocation, we used several algorithmic clustering techniques. The 112 models examined in detail were "counter clustered", i.e. the direction of the analysis was reversed: the purpose of the manual clustering had been to find groups of models sharing a primary focus and to then examine the differences in complexity of these groups. During the algorithmic clustering process, we grouped models according to their complexity properties and the resulting clusters were compared regarding their focus. The methods chosen were *k-pod clustering*, *hierarchical clustering* and *multiple correspondence analysis* (MCA) combined with *k-means* clustering. For all methods that required choosing the number of clusters, we set it to $k = 4$ in order to test whether the manual clustering's results could be replicated.

k-pod is a clustering method developed by Chi et al. [35] that is derived from *k-means*. It combines the ease-of-use of the popular *k-means* method with the capability to deal with missing data. Generally, clustering algorithms require datasets to be complete [36]. However, several strategies to adapt the dataset exist: (1) the analysis can be restricted to either only those items or those attributes that are complete or (2) the missing data can be imputed (estimated) [37]. Reducing the dataset was infeasible for our dataset, since there was missing data for all models and all attributes, so we would have lost too much information. The second strategy decreases the clustering results' quality, so working with a clustering algorithm that is capable of dealing with missing data points is preferable. Chi et al. [35] tested *k-pod* against other strategies to deal with missing data points, demonstrating its superiority with regards to clustering accuracy. Hence, we initialized *k-pod* ten times and chose the results of the run that produced the lowest *sum of squared distances* (SSD) [38] for our analysis. In order to examine the quality of the clustering results, we compared SSD and silhouette score [39] results for two to ten clusters.

Like *k-means*, *k-pod* is based on *Euclidean distance* as a measure of (dis)similarity. This can negatively affect clustering results when working with binary data. Generally, both binary data [40] and high-dimensional data [41] is difficult to cluster, so few promising methods for data that possesses both of these qualities exist. One of them is hierarchical clustering, which is compatible with a variety of distance measures. Choi et al. [42] compare 76 measures for binary data. We chose normalized *Hamming distance* to measure the similarity between the models. Being a symmetrical dissimilarity measure, *Hamming distance* assumes that zeroes and ones (i.e. occurrences and non-occurrences of attributes) carry equal meaning. Since the modelers made conscious choices out of a number of limited options (e.g. which demand sectors to include), we opted for a symmetrical measure.

There are different types of hierarchical clustering algorithms. Agglomerative algorithms start by creating a cluster for each of the objects that are to be clustered [41]. The distances (i.e. dissimilarities) between the clusters are calculated and the two clusters most similar to each other are merged. This process is repeated until the number of clusters pre-specified by the user is reached. Divisive algorithms start by creating one cluster for all objects, which is then divided into smaller clusters [41]. The algorithm we used was an agglomerative one.

The rule according to which two clusters are merged by an agglomerative algorithm is called *linkage method*. For example, the algorithm can apply the chosen distance measures to all elements of the two clusters to calculate their distance (average linkage), only consider the elements that are farthest from each other between the two clusters (complete linkage) or those that are closest (single linkage). The linkage method chosen was average linkage in order to consider the clusters'
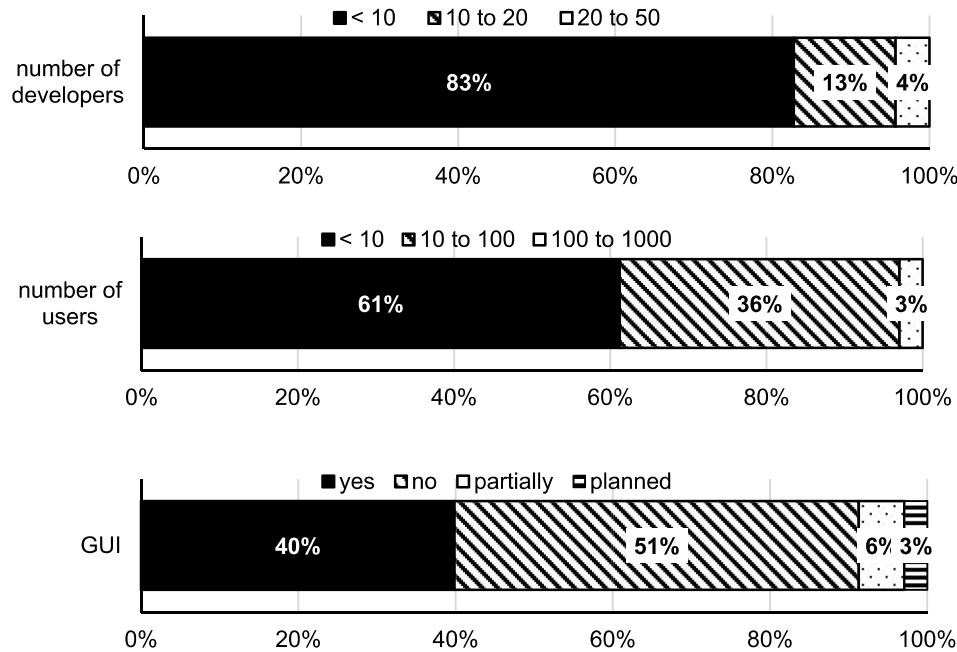
Fig. 1. Number of model developers and users and share of models possessing a GUI.

whole makeup and minimize misallocations. Missing data was imputed using column means [34].

Since neither *k-pod* nor *hierarchical clustering* possess means to specifically address high-dimensional data, we chose a tool of dimension reduction as our third clustering method. Performing a MCA on the data allowed us to both reduce its dimensionality and transcribe it into a non-binary format [43]. In an MCA, an *eigenvalue problem* is solved, using the first *n eigenvectors* as a new coordinate system to describe the data. The number of *eigenvectors* used is picked depending on the level of variance to be explained [44]. Choosing a higher number of *eigenvectors* allows covering more of the data's variance but limits the dimension-reducing effect of the MCA. In order to cover 95% of the variance in our data, we performed the MCA with 50 *eigenvectors*. The transcribed data was then clustered using *k-means*. Missing data was imputed using column means.

5. Results

The manual clustering process resulted in 10 clusters of main *foci*.

Table 3
Number of models in the clusters.

Cluster	Description	No. of models	Share of total models
Electrical grids	Analyzing and optimizing electrical grids	39	27%
Future energy systems	Scenario analysis for energy system transformations	35	24%
Unit commitment	Power plant deployment planning	21	14%
Policy assessment	Scenario analysis for evaluating the effects of energy-related policies	17	12%
Decentralized energy systems	Optimizing the planning, extension and operation of decentral generation units	8	6%
Societal influences	Analyzing the socio-economic framework of energy systems	8	6%
Building sector	Optimizing the energy supply for building sector	5	3%
Data processing	Data processing methods for energy system models	3	2%
Other	Varying foci, none of the above	9	6%

Table 3 shows the clusters sorted by the number of models assigned to them.[1] The four largest clusters are noticeably larger than the other clusters. Together, they contain 112 out of the 145 models considered (77%). These were the four clusters whose complexity properties were examined in detail.

The cluster named "*electrical grids*" contains models that are used to analyze electrical grids and optimize their operation. Among other things, they model load flows or the effects of battery storage and decentral electricity generation on the grid. Models in the cluster "*future energy systems*" are used for scenario analyses with regard to energy system transitions. They focus on future technologies, CO_2 emissions and energy demands. Models in the cluster "*unit commitment*" are used to obtain power plant deployment schedules. Some of the models also include flexibility options and investment decisions. The last of the four big clusters, "*policy assessment*", contains models used to assess the effects of energy-related policies.

5.1. Complexity properties of the clusters

Since a complete description of the complexity properties of the clusters would go beyond the scope of this paper, only selected examples based on the main complexity drivers as introduced in section 2.2 will be presented: the mathematical modeling approach, the temporal horizon and resolution and the modelled system scope.

5.1.1. Modeling approach

Fig. 2 shows the share of modeling approaches in each cluster. Additionally, the leftmost group of columns shows the distribution of approaches across all models, including the small clusters not explicitly pictured.

With 67% and 74%, respectively, *linear programming* (LP) is the most popular modeling approach for unit commitment models and models of future energy systems. In these two clusters, *mixed-integer linear programming* (MILP) is the second-most often used approach. Models used for the analysis of electrical grids and for policy assessment, however,

---

[1] In principle, a model could be part of different clusters if it covers different foci. However, for reasons of simplicity we assigned each model to the best fitting cluster, only.
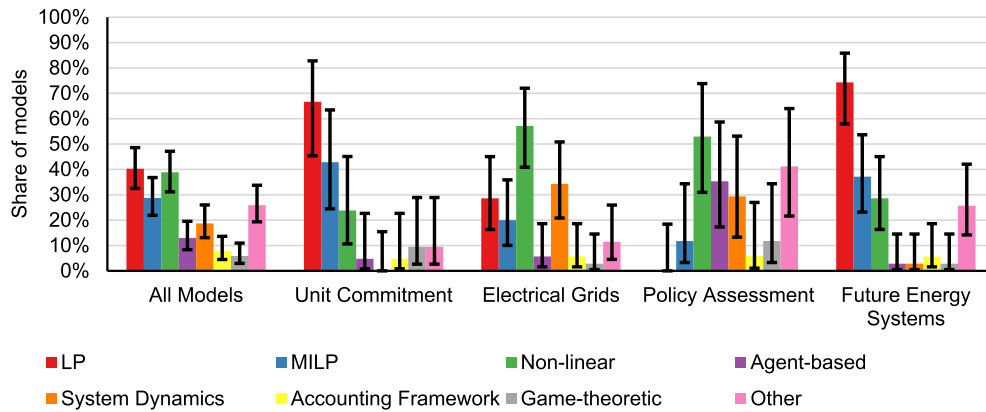
Fig. 2. Modeling approaches with Wilson Score intervals (95% significance level).

predominantly use non-linear approaches, including *"System Dynamics"*. *System dynamics* is a multi-disciplinary modeling approach capable of examining problems that are non-linear and subject to feedback loops [45]. It has found use in a variety of disciplines, including economics, energy and environment modeling and policy analysis [46]. Given that it is based on system and control theory [47], it stands to reason that this approach is well-suited to analyzing electrical grids.

The *Policy assessment cluster* stands out as the only cluster lacking LP models and having a high share (41%) of *"other"* responses. Most of its models are non-linear in nature, including *agent-based* and *system dynamics approaches*. Since the questionnaire allowed respondents to select any of the categories concurrently, the results may over-present the unspecific *"non-linear"* category. The comments explaining the *"other"* responses either offer additional information regarding the categories selected or reveal that the corresponding models' approaches are combinations of several modeling techniques, making them difficult to classify. With one exception, the comments given by creators of *future energy systems* models are of an explanatory nature. This indicates that the modelers felt that their modeling approach could be classified within the categories available but wanted to give additional information.

The black bars in the diagram show the 95% confidence intervals calculated. Their meaningfulness varies depending on the cluster. For example, the intervals show that LPs are the dominant modeling approach in the *future energy systems cluster* on the chosen confidence level. It is not possible to make that same statement for the other clusters, however, since the confidence intervals overlap. This is part of a general trend that exists across the whole database. For this reason, further commentary on the confidence intervals will be given referring to the data as a whole in section 6.

5.1.2. Temporal horizon and resolution

Fig. 3 shows the models' temporal horizon and resolution. Generally, temporal resolutions and horizons are balanced, keeping the number of time steps to calculate at a minimum. The vast majority of unit commitment models are run using an hourly resolution in combination with a temporal horizon of one year. Policy Assessment models lie at the opposite end of this spectrum, since they tend to be long-term models (76%) with only a yearly temporal resolution. Some of these models offer horizons of up to 100 years. Models within the *future energy systems cluster* are similar to *policy assessment models* in that they are often long-term models as well. However, the *future energy systems* models use significantly higher resolutions, often being run with hourly resolutions. The high number of replies in the *"other"* category is due to these models often supporting user-specified resolutions or using a fixed number of time slices per year.

Models of electrical grids use a variety of resolutions and horizons. Most notable is the high number of short-term models with a horizon of less than one year as well as the high number of replies (61%) indicating
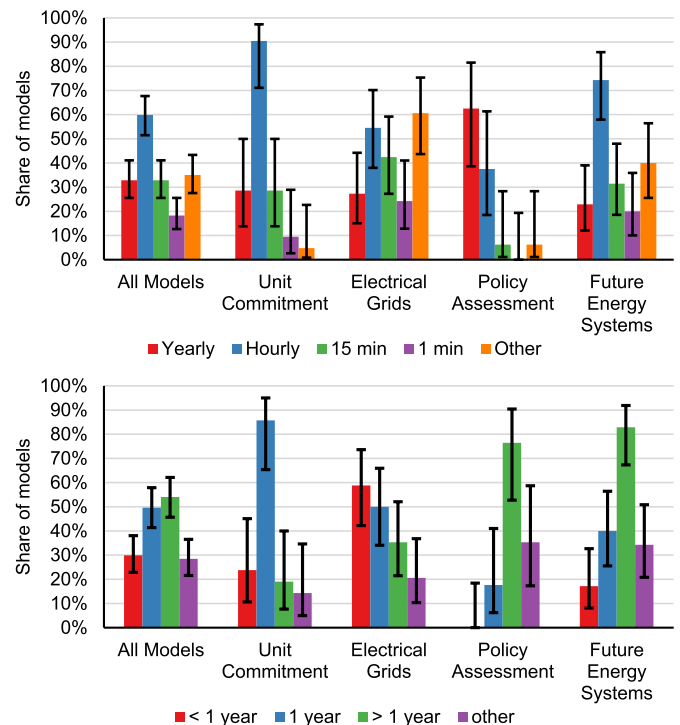

Fig. 3. Temporal horizon and resolution of the models with Wilson Score intervals (95% significance level).

"other" resolutions. These replies fall into three groups: 24% of all electrical grid models use very high resolutions smaller than seconds, 15% support user-specified resolutions and 6% of all respondents indicated that their models do not have a temporal resolution. The remaining 16% of respondents indicating *"other"* did not give further explanations.

5.1.3. System scope

As shown in Fig. 4, across all clusters, the electricity sector is modelled most frequently. All of the models in the *unit commitment cluster* model electricity, while 62% of these models also model heat. The *electrical grid cluster* is clearly focused on the electricity sector, with none of the other sectors being included in more than 20% of the models. *Policy assessment* and *future energy systems* models display the greatest breadth of sectoral representation, with at least half of the models in both clusters including the electricity, hear, liquid fuels and gas sectors.
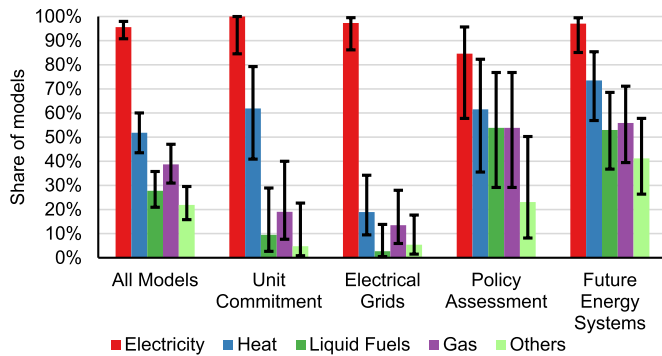
Fig. 4. Energy generation sectors modelled with Wilson Score intervals (95% significance level).

5.2. Complexity profiles

Based on the analysis of the clusters' complexity properties that has been presented in part above, we derived complexity profiles for the four clusters examined. The profiles describe the areas of high and low complexity of the clusters' models.

5.2.1. Unit commitment

The models in the unit commitment cluster are predominantly characterized by mathematically rather less complex approaches. Most of them are LP models – or less often MILP models – that focus on modeling the cost of energy supply. They usually operate with an hourly resolution and a temporal horizon of one year. Like the models of all clusters, they cover a broad variety of spatial scopes. Most models focus on the national level or on the level of individual power plants. The complexity of the system modelled varies. Their modeling of energy supply and demand sectors is on par with that of other clusters' models with a focus on the electricity sector. Apart from electricity, only the heat sector appears to be important for current *unit commitment* models as less than 20% of the models depict one of the other sectors. Models in this cluster tend to include a broad variety of energy carriers and electricity generation technologies. Overall, transmission capacities as well

as storage technologies are examined more frequently by models in the *unit commitment cluster* compared to the other clusters.

The modelers' comments in the data indicate that models in the *unit commitment cluster* are predominantly used to solve (power plant) dispatch and investment problems. This is in line with their complexity profile, as evidenced by this cluster's focus on the electricity sector and power plant modeling.

The fact that heat is the second-most relevant generation sector in this cluster is possibly explained by heat being a by-product of thermal power plants [48]. Other sectors are excluded from analysis, indicating that covering electricity generation in great depth and breadth takes precedence over examining sector coupling potentials. It is also noteworthy that *unit commitment* models often model storage technologies. Pumped storage hydroelectric plants are regarded as a storage technology with great future potential [48], which might explain the fact that all of the models in the *unit commitment cluster* include water-based generation technologies. This conclusion is supported by the fact that these models often include flexibility options, such as demand side management and storage.

Despite their similarities in modeling approaches and purpose of the analysis, *unit commitment* models possess a far shorter temporal horizon than *future energy systems* models. This indicates that models in the *unit commitment cluster* are used for short-term power plant operation planning rather than exploration of possible energy system transformation paths. Their focus is the optimization of an existing generation portfolio including relevant investment decisions in the near future.

All findings mentioned above are summarized in Fig. 5.

5.2.2. Electrical grids

The *electrical grids cluster* is characterized by high mathematical complexity. Its models focus on examining technical parameters requiring non-linear approaches. The models' temporal resolution varies from hourly to frictions of seconds. Their temporal horizon is often limited to the short term, which is likely an attempt to limit the time steps that have to be calculated. Regarding spatial complexity, this cluster differs from the others in the sense that individual power plants and grid areas of transmission system operators are modelled particularly often.

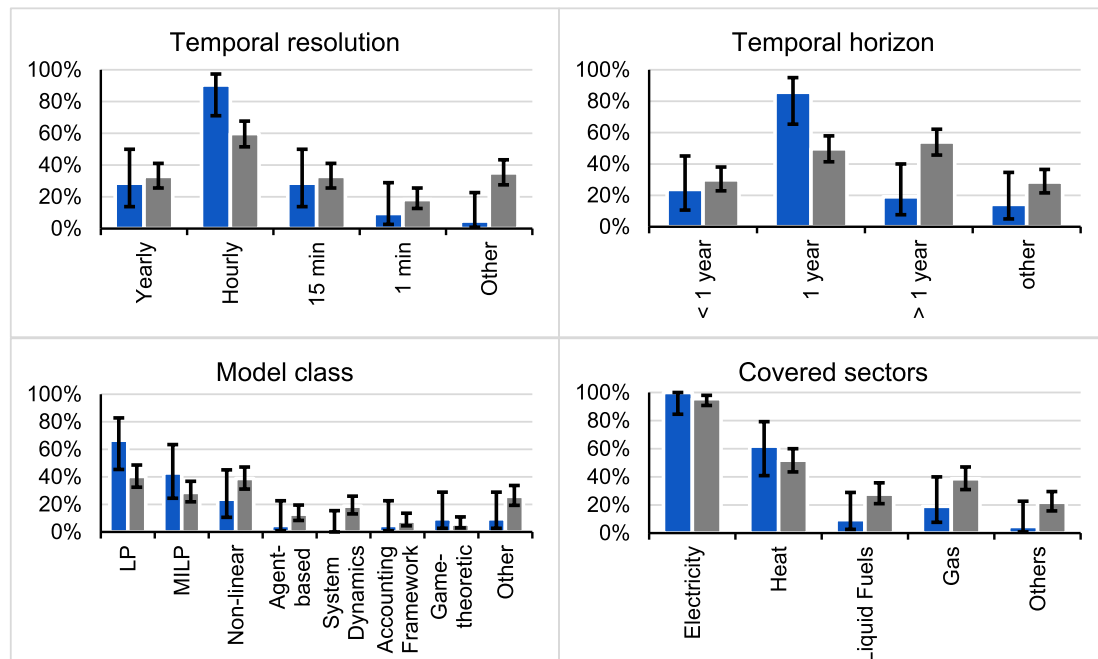Their high temporal and mathematical complexity is counter-

Fig. 5. Comparison of unit commitment models (left blue column) and all models (right grey column) with Wilson Score intervals (95% significance level).

balanced by limiting the complexity of the content modelled to highly grid-relevant elements. On the supply side, there is a focus on the electricity sector. On the demand side, the industry and trade sectors are more often included (71% and 65%, respectively) than households and transport (48% and 35%). With the electrification of transport and its implications on electrical grids currently being a topic of high interest, the transport sector being among the rarely included sectors is surprising. The industry sector being more often included in *electrical grid* models than households is expected, however. This is due to the fact that demand side management, another highly grid-relevant topic, is seen to have more potential in industrial applications than in the household context [49].

As opposed to the other clusters, models in the *electrical grids cluster* examine load flows to a greater extent than transmission capacities. Half of all models in this cluster include direct current load flows. Some of the model authors indicated their models examine high voltage direct current, which is regarded as a necessary technology for a pan-national *electricity grid* [50]. In addition, it offers possibilities of improved load distribution and feed-in for large-scale renewables (such as for offshore wind parks) [51]. This indicates that at least part of the models are concerned with analyzing future pan-national grids.

Again, Fig. 6 summarizes our findings.

5.2.3. Policy assessment

The models in the *policy assessment cluster* use non-linear modeling approaches, such as *system dynamics* and *agent-based modeling*. While there is a focus on economic effects, these models examine all three target dimensions of energy policy (economic efficiency, environmental sustainability and security of supply, cf. [56]). Most *policy assessment* models possess a long temporal horizon combined with a low, often yearly, resolution. Regarding spatial complexity, all *policy assessment* models cover national states, while none depict individual power plants.

The fact that they often use agent-based modeling indicates that agents' reactions are the focus of the analysis. This is supported by the fact that the *policy assessment cluster* possesses a high number of market models on the one hand and households being the most often modelled demand sector on the other hand.

The content modelled is less broad than in other clusters, in particular with regard to detailed grid modeling. *Policy assessment* models

examine gasoline as well as diesel fuels far more often and the primary energy carriers hydrogen and uranium far less often than other models. This indicates that policy modelers limit complexity by focusing on those parts of the energy system that are deemed of high societal importance: hydrogen research funding is considerably lower than that of other renewable energy sources (see Ref. [52]), while uranium has lost importance as an energy carrier due to ongoing nuclear phase-outs. On the other hand, liquid fuel prices are of high importance across society and affect electric vehicles' adoption [53].

The complexity profile of policy assessment models is illustrated in Fig. 7.

With respect to its modeling approaches, the *future energy systems cluster* is similar to the *unit commitment cluster*. Models within this cluster use linear approaches and examine economic variables, mostly. The models in this cluster use long temporal horizons, frequently covering several decades. Still, an hourly resolution is the most common choice (74%), making these models highly temporally complex. Some models use custom resolutions, such as a set number of time slices per year. This indicates that modelers try to limit the number of time steps to be calculated. With regard to spatial complexity, *future energy systems* models examine households less often and power plants more often than other models. This distinguishes them from *policy assessment* models in particular and is verifiable through the 95% confidence intervals.

Apart from their modeling approach, there are further similarities between *future energy system models* and *unit commitment* models. Just like the latter, *future energy systems* models often require solving both dispatch and investment problems, as evidenced by the modelers' comments. Thus, one can conclude that as opposed to policy models, *future energy systems* models examine not agents' actions but technological and financial feasibility. In order to do so, they include a broad variety of technologies, resulting in high modeling content complexity. In all subcategories of this complexity dimension, their inclusion of features is high, often on par or slightly above that of *unit commitment* models and considerably exceeding that of the other two clusters. This distinguishes *future energy systems* models further from *policy assessment* models.

Fig. 8 summarizes our findings regarding the complexity dimension of future energy systems in comparison to other model types.
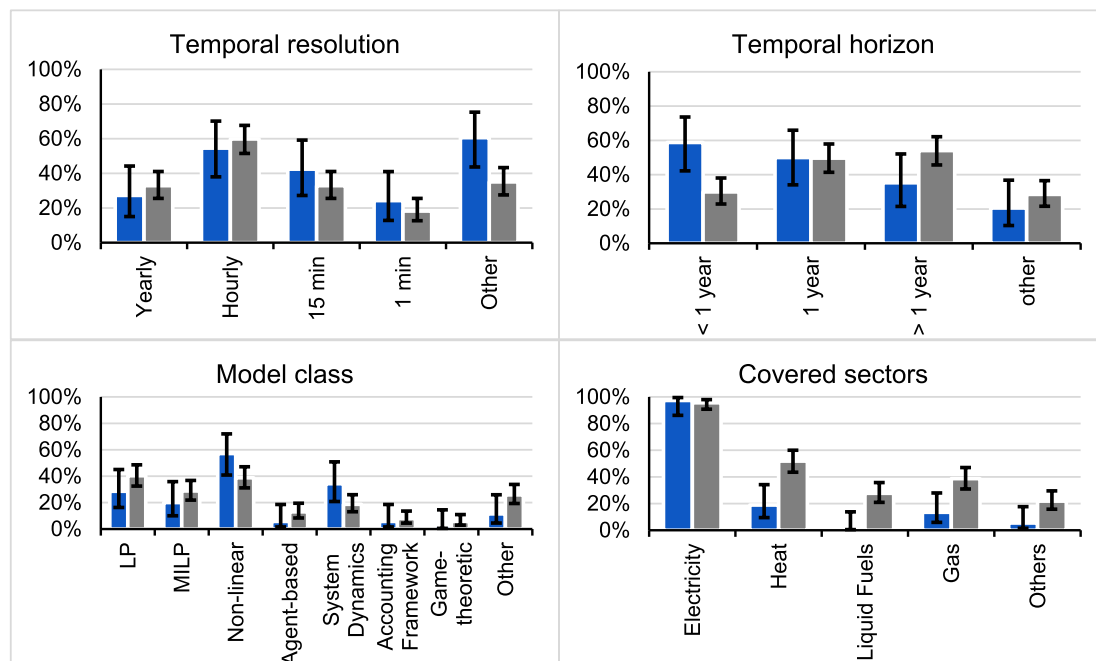

Fig. 6. Comparison of electical grid models (left blue column) and all models (right grey column) with Wilson Score intervals (95% significance level).
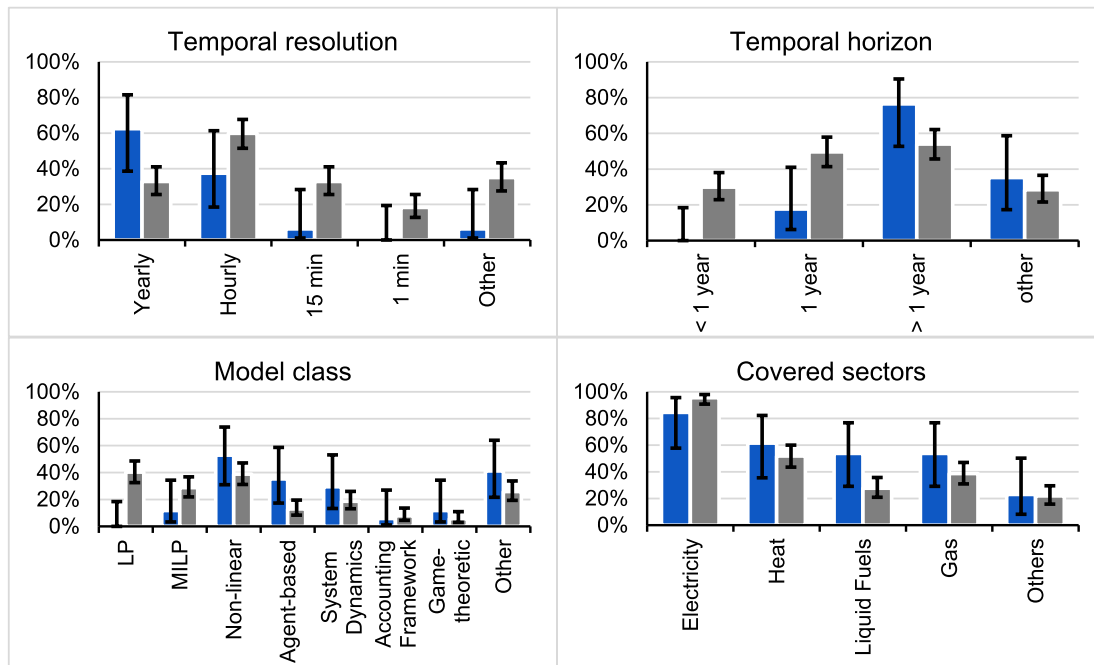
Fig. 7. Comparison of policy assessment models (left blue column) and all models (right grey column) with Wilson Score intervals (95% significance level)Future energy systems.
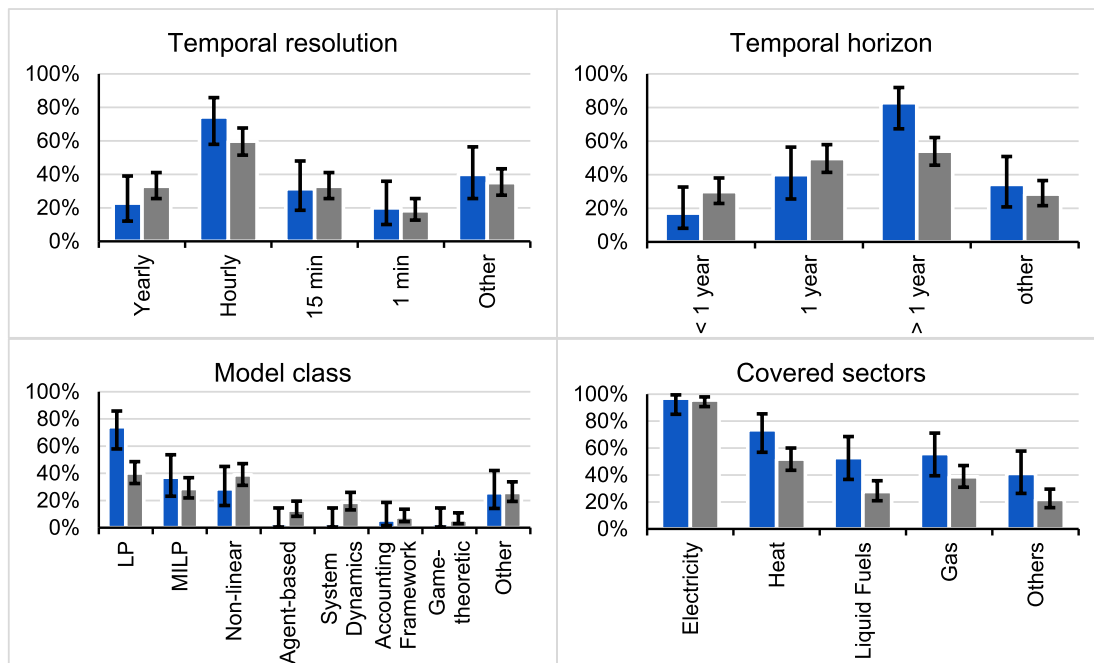

Fig. 8. Comparison of policy assessment models (left blue column) and all models (right grey column) with Wilson Score intervals (95% significance level).

5.3. Algorithmic clustering

Having examined the manually identified clusters, we used several clustering algorithms in order to test the validity of our clusters. For reasons of brevity, only the results obtained using *k-pod* and *hierarchical clustering* are explained in detail. The results of the MCA will be summarized in section 6 (discussion).

5.3.1. K-pod

We compared the cluster allocation of the *k-pod* run that generated the lowest SSD to the allocation achieved through the manual clustering

process. Table 4 shows the models' allocation ordered by manual clusters. The clusters created by *k-pod* were given the letters A to D.

The models in the cluster *electrical grids* and, to a lesser extent, those in the cluster *unit comment* were mostly allocated to one cluster by *k-pod*. This is not true for *policy assessment* and *future energy systems* models, which the algorithm allocated into its four clusters very evenly. Comparing the complexity properties of the *k-pod* clusters to the manual clusters', the manual ones are far more clearly distinguishable. It was not possible to generate complexity profiles based on the *k-pod* clusters.

In order to evaluate the number of clusters chosen, we compared SSD values and silhouette scores for up to ten clusters. As shown in Fig. 9, the

Table 4
Comparison of manual clusters and K-Pod clusters.

Cluster (manual)	Cluster k-pod	Number of models	Share
Unit commitment (21 models)	A	6	29%
	B	3	14%
	C	12	57%
	D	0	0%
Electrical grids (39 models)	A	5	13%
	B	30	77%
	C	3	8%
	D	1	3%
Policy assessment (17 models)	A	0	0%
	B	7	41%
	C	4	24%
	D	6	35%
Future energy systems (35 models)	A	11	31%
	B	6	17%
	C	9	26%
	D	9	26%

SSD curve is flat without a noticeable "elbow". The highest silhouette score sil_{max} occurs for two clusters and is very small ($sil_{max} = 0.185$). We also applied the *jump method* (results not shown), finding that its results varied heavily by run. Sugar and James [54] recommend selecting an exponent smaller than their suggestion of $Y = \frac{m}{2}$ for high-dimensional data. We tested the values 5, 10, 30 and 50 ($\approx m/2$) for Y and ran several iterations for up to 10 clusters. Depending on the run, different cluster numbers appear to be optimal. In three out of four runs, the maximum cluster number chosen (ten) is among the ones that appear best suited. In an additional run with up to 20 clusters, 20 appeared as the optimal cluster number. This indicates that applied to our data, the jump method tends to evaluate the maximum cluster number as one of the best-fitting numbers, regardless of which cluster numbers are considered.

5.3.2. Hierarchical clustering

Table 5 shows the clusters created by the *hierarchical clustering* algorithm. The results are ordered by the manual clusters. At first glance, the results seem to validate the cluster allocation that has been manually determined. In all clusters but *electrical grids*, the models are overwhelmingly allocated to one cluster. However, in each case, this is algorithmic cluster C. An examination of the algorithmic clusters' sizes (right in the Figure) reveals that cluster C contains 80 of the 112 models.

6. Discussion

In this part, the results of the manual and algorithmic clustering processes will be discussed.

6.1. Manual clustering

The results of the manual clustering showed a connection between the purpose of a model and its complexity. We found four clusters with clearly distinguishable purposes and complexity profiles. As we demonstrated a substantial relation between the main thematic focus of a model (i.e. the research questions to be answered using it) and its main complexity attributes (the so-called complexity profiles), we can derive that conscious decisions have been made by the modelers. A more detailed depiction with regard to one complexity driver is compensated by less accuracy with regards to other complexity drivers. Based on this, we conclude that modelers ***allocate complexity*** according to their priorities: depending on a model's focus and purpose, modelers prioritize different dimensions of complexity by conducting trade-offs.

The differences between the clusters vary depending on the complexity property examined. There are only minor differences with regard to spatial complexity, for example. This may be due to the format of the MODEX questionnaire, which did not distinguish spatial coverage from spatial resolution. In most properties, there are substantially more pronounced differences between the clusters. This indicates that (1) there is indeed a broad variety of models and that (2) the clusters capture the differences between different model types well.

It is noteworthy that the intra-cluster variance of the data varies by property as well. While in most properties the profiles are clear-cut, in

Table 5
Clusters as created by Hierarchical clustering.

Cluster (manual)	Cluster Hierarchical	Number of models	Share
Unit commitment	A	0	0%
	B	0	0%
	C	21	100%
	D	0	0%
Electrical grids	A	3	8%
	B	16	41%
	C	17	44%
	D	3	8%
Policy assessment	A	0	0%
	B	3	18%
	C	14	82%
	D	0	0%
Future energy systems	A	1	3%
	B	5	13%
	C	28	72%
	D	1	3%

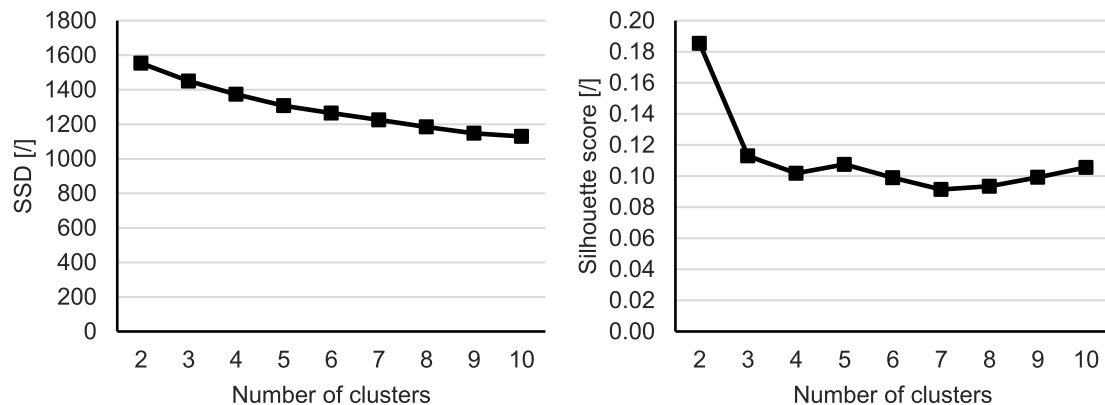| Cluster (hierarchical)	Number of models	Share
A	4	4%
B	24	21%
C	80	71%
D	4	4%

Fig. 9. SSD (left) and silhouette score (right) for K-Pod.

some cases the spread between the models within one cluster is rather large. This is the case for the *electrical grids* cluster in particular. For example, while this cluster's temporal horizon and its temporal resolution are easily distinguished from the other clusters', they are distributed more evenly (see Fig. 3). This offers the conclusion that the *electrical grids* cluster includes a wider range of models, which differ in their complexity properties.

The results of the manual clustering process are subjective to a certain extent. This is not only due to the clustering process applied itself, but also because the data it is based on, i.e. the modelers' statements on their models' purpose, is sometimes ambiguous. In particular when models are created with several purposes in mind or are used to answer research questions that they had not been designed for, it is not easy to make clear-cut distinctions.

The complexity profiles described above are based on the differences between the percentages we calculated regarding the clusters' complexity properties. These differences can only be statistically validated on a 95% confidence level in some cases. As Galvin [34] explains, there are three factors that influence the size of *Wilson Score Intervals*:

- The sample size n
- The confidence level p
- The value to be validated

An increase in sample size decreases the size of the confidence interval. That is why the intervals of the percentages calculated for all 145 models are smaller than the clusters', while the smallest cluster's (i.e. *policy assessment*) intervals are the largest ones. A bigger sample size might thus help in statistically validating the differences between the clusters. However, this is hard to realize. To our knowledge, the MODEX fact sheets constitute the biggest concerted effort to comprehensively gather model data to date.

The confidence level chosen was 95%. It is intuitively understood that selecting a lower confidence level (i.e. a lower probability that the true value lies in the confidence interval) results in smaller confidence intervals. That is why we tested 90% confidence levels for some of complexity properties. Since this brought only small improvements, we did not carry out a full analysis with 90% intervals.

Finally, the value to be validated itself has an influence on its confidence interval. Values close to zero and one possess smaller intervals than those in the middle of the percentage scale. This cannot be controlled during confidence interval calculation but influences whether differences can be validated. Such values that lie close to the middle of the scale are disproportionally difficult to verify. This is the case for a majority of the percentages calculated, since there are few complexity properties that are present or lacking in a large part of the models.

6.2. Algorithmic clustering

The clustering algorithms' results differ to great extent. This was to be expected as clustering represents an explorative technique. As a method of unsupervised learning, there is no correct or ideal result to judge an algorithm's output against. Thus, a single algorithm's clustering results is to be regarded as one possible way of structuring the data rather than the data's "real" structure. The differences in the cluster allocations are therefore a result of the clustering algorithms' different approaches and of the data's "clusterability". Both a high number of dimensions and a binary data format complicate clustering efforts.

6.2.1. K-pod

The *k-pod* results indicate that the Unit Commitment and Electrical Grids clusters are clearly distinguishable, whereas the other two clusters are not. This might indicate 4 being a poorly-chosen cluster number. It is possible that the two clusters that could not be easily distinguished in fact consist of more clusters.

This interpretation, however, seems implausible given that k-pod's

results seem of low quality. The silhouette value indicating two clusters can either be due to two clusters describing the data best, due to there being little structure in the data or due to the silhouette value being ill-fitted to the data. The latter is likely due to the data being high-dimensional. With high-dimensional data, differences between distances lose meaning [55]. Since the silhouette value is calculated using these differences, it is unreliable for high-dimensional data. This conclusion is supported by the low silhouette values overall.

The flat SSD curve does not allow conclusions regarding the number of clusters. Thus, either there is little structure in the data or the elbow method fails to detect the structure. We used the *jump method* to distinguish between these two cases. The *jump method* showed clear differences between the cluster numbers tested, but which ones were indicated as well-suited and ill-suited, respectively, was not consistent over several runs. This indicates that the allocation of the models to clusters was not consistent either.

Given that *k-means* uses *Euclidean distances* to calculate the differences between the objects to be clustered, this is plausible. *Euclidean distance* is best used with metric data. While it is possible to use it with binary data, it can impair the meaningfulness of the cluster allocation. Although the missing data was filled in with metric data, it is possible that this has happened. Thus, it is likely that the cluster allocation is a result of *k-means'* properties rather than a reflection of the data's actual structure.

6.2.2. Hierarchical clustering

As opposed to *k-means*, *hierarchical clustering* algorithms are compatible with a broad range of pairwise distance measures. This makes them a likely choice for binary data. However, the results of this clustering method again display the algorithm's properties as well rather than a structure underlying the data. The cluster sizes differ vastly, which is likely a result of the choice of algorithm, since agglomerative clustering tends to result in uneven cluster sizes.

We examined whether this was a result of the *linkage method* chosen and clustered the data again using complete linkage (i.e. the distances between clusters were calculated using those elements in the clusters that were farthest from each other). This resulted in more even cluster sizes. However, since complete linkage determines two clusters' similarity by only one element in each cluster, this can lead to misallocations. That is why we did not further analyze this cluster structure.

6.2.3. Dimension reduction

Reducing the number of dimensions and converting the data into a metric format through an MCA was expected to result in clusters that differ from those determined using *k-pod*. In part, this was true. However, the models in one of the four manually determined clusters (i.e. *unit commitment*), were allocated to the exact same clusters through both *k-pod* and *k-means* after performing the MCA. Examining some of the clusters' complexity properties, they were less clearly distinguishable than the clusters determined manually. Since a high number of dimensions had to be used in order to cover 95% of the data's variance, the dimension reduction effect of the MCA was small. It is likely that the remaining high dimensionality impaired the clustering algorithm's results.

6.3. Comparison of clustering approaches

While the results of the algorithmic clustering approaches remained inconclusive to a certain extent, the manual clustering approach led to clearly distinguishable clusters with separate complexity profiles. We thus conclude that due to the binary nature and the high dimensionality of the data under investigation, manual clustering allows for better insights. Therefore, we derived the complexity profiles based on the manual clustering method and concluded the existence of trade-offs between different dimensions of complexity based on these results.

7. Conclusion

Using data gathered in the MODEX fact sheets, we conducted both manual and algorithmic clustering to carve-out distinguishable model types. Based on this categorization, we investigated the models' key attributes with regards to complexity.

The manual clustering procedure resulted in four distinguishable clusters of different thematic foci, possessing well-defined complexity profiles. The results indicate that there is indeed a relationship between the research questions a model is supposed to answer and its complexity. We thus conclude that modelers **allocate complexity** by prioritizing those aspects that are deemed most important for answering the desired research questions (i.e. a *trade-off* between different drivers of model complexity are made). These aspects then represent the real system's qualities in greater detail, while others are regarded as less important for the model's main focus.

The algorithmic clustering procedures intended to verify the clusters resulted in different cluster structures depending on the algorithm. Since clustering is an explorative technique, a clustering algorithm can only suggest one possible structure that underlies the data. A single algorithm's results are not to be interpreted as the "real" structure of the data. As the manual clustering process did result in clearly distinguishable complexity profiles, we concluded that there is indeed a structure present in the data that the algorithms failed to capture.

The existence of a **complexity allocations** shows that there is awareness for complexity in the modeling community. Our findings indicate first efforts towards limiting model complexity by modeling at a high level of detail only where this is needed. However, this is only a first step towards comprehensive complexity management. If there is to be a comprehensive approach that allows fine-tuning the trade-off between models' level of detail and the resulting complexity in various dimensions, more discussion on guidelines and best practices is needed. To this end, the current push towards a more enhanced complexity management in energy system analysis can be claimed a necessary step.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Appendix A

Attributes	Possible replies (if applicable)	Used for clustering?
1. General information		
Model name	Open question	Not used
Acronym	Open question	Not used
Institutions	Open question	Not used
Authors	Open question	Not used
Current contact person	Open question	Not used
Contact email	Open question	Not used
Website	Open question	Not used
Support/community/forum	Open question	Not used
User documentation	Open question	Not used
Code documentation	Open question	Not used
Documentation quality	Open question	Not used
Source of funding	Open question	Not used
Number of devolopers	Open question	Not used
Number of users	Open question	Not used
Citation reference	Open question	Not used
Citation DOI	Open question	Not used
References to reports produced using the model	Open question	Not used
Larger scale usage	Open question	Not used
2. Thematic focus		
Primary purpose	Open question	Manual
Primary outputs	Open question	Manual
Example research questions	Open question	Manual
Methodical focus	Open question	Manual
3. Licensing and openness		
Open source	Open question	Not used
Going to open up	Open question	Not used
Costs	Open question	Not used
License	Open question	Not used
Source code available	Open question	Not used
GitHub	Open question	Not used
Link to source code	Open question	Not used
Data provided	Open question	Not used
Cooperative programming	Open question	Not used
I agree to publish this factsheet on the open energy platform (OEP)	Open question	Not used
4. Programming		
Framework	Open question	Not used
Framework name	Open question	Not used
Software	Open question	Not used
Modeling software	Open question	Not used
Internal data processing software	Open question	Not used

(*continued*)

Attributes	Possible replies (if applicable)	Used for clustering?
External optimizer	Open question	Not used
Additional software	Open question	Not used
GUI	Open question	Not used
References	Open question	Not used
Typical computation time	Open question	Not used
Typical computation hardware	Open question	Not used
Technical data anchored in the model	Open question	Not used
Model Integration	Open question	Not used
Interfaces	Open question	Not used
Model file format	Open question	Not used
Model input	Open question	Not used
Model output	Open question	Not used
Integrating models	Open question	Not used
Integrated models	Open question	Not used
5. Complexity properties		
Modelled energy sectors	Electricity, heat, liquid fuels, gas, others	Algorithmic
Modelled demand sectors	Households, industry, commercial sector, transport	Algorithmic
Modelled energy carriers (primary energy carriers)	Gas: natural gas, biogas, hydrogen	Algorithmic
	Liquid: petrol, diesel, ethanol	
	Solid: hard coal, lignite, uranium	
	Renewables: biomass, sun, wind, hydro, geothermal Heat	
Modelled technologies: components for generation or conversion	Renewables: PV, wind, hydro, biomass, biogas, solar thermal, other renewables (+comment)	Algorithmic
	Conventional: gas, coal, oil, liquid fuels, nuclear	
	CHP	
Modelled technologies: components for transfer, infrastructure or grid	Electricity: distribution, transition	Algorithmic
	Gas: distribution, transition	
	Heat: distribution, transition	
Modelled properties electrical grid	AC load flow, DC load flow, net transfer capacities	Algorithmic
Modelled technologies: components for electricity storages	Battery, kinetic, compressed air, pump hydro, chemical	Algorithmic
Modelled technologies: components for other storages	Heat, gas	Algorithmic
User behavior	Open question	Algorithmic
User behavior and demand side management	Open question	Algorithmic
Changes in efficiency	Open question	Algorithmic
Market models	Open question	Algorithmic
Spatial and temporal model properties	Open question	Algorithmic
Geographical coverage/Geographic (spatial) resolution of the model	Global, continents, national states, TSO regions, federal states, regions, NUTS 3, municipalities, districts, households, power stations, others, comments	Algorithmic

References

[1] S. Pfenninger, L. Hirth, I. Schlecht, E. Schmid, F. Wiese, T. Brown, et al., Opening the black box of energy modelling: strategies and lessons learned, Energy Strategy Reviews 19 (2018) 63–71, https://doi.org/10.1016/j.esr.2017.12.002.

[2] IEA, IEA-ETSAP Community, 2020 (accessed March 27, 2020), https://iea-etsap.org/.

[3] C.S.E. Bale, L. Varga, T.J. Foxon, Energy and complexity: new ways forward, Appl. Energy 138 (2015) 150–159, https://doi.org/10.1016/j.apenergy.2014.10.057.

[4] N. van Beeck, Classification of energy models, FEW Research Memorandum 77 (1999) 26.

[5] S. Winkelmüller, Optimierung der Nachfrage- und Erzeugungsstruktur kommunaler Energiesysteme am Beispiel von Wien (Optimisation of the demand and generation structure of municipal energy systems using Vienna as an example), Dissertation, Universität Augsburg, Germany, 2006.

[6] H.-K. Ringkjøb, P.M. Haugan, I.M. Solbrekke, A review of modelling tools for energy and electricity systems with large shares of variable renewables, Renew. Sustain. Energy Rev. 96 (2018) 440–459, https://doi.org/10.1016/j.rser.2018.08.002.

[7] D. Möst, D. Keles, A survey of stochastic modelling approaches for liberalised electricity markets, Eur. J. Oper. Res. 207 (2010) 543–556, https://doi.org/10.1016/j.ejor.2009.11.007.

[8] G. Savvidis, K. Siala, C. Weissbart, L. Schmidt, F. Borggrefe, S. Kumar, et al., The gap between energy policy challenges and model capabilities, Energy Pol. 125 (2019) 503–520, https://doi.org/10.1016/j.enpol.2018.10.033.

[9] R. Weijermars, P. Taylor, O. Bahn, S.R. Das, Y.-M. Wei, Review of models and actors in energy mix optimization–can leader visions and decisions align with optimum model strategies for our future energy systems? Energy Strategy Reviews 1 (2012) 5–18.

[10] S. Pfenninger, A. Hawkes, J. Keirstead, Energy systems modeling for twenty-first century energy challenges, Renew. Sustain. Energy Rev. 33 (2014) 74–86, https://doi.org/10.1016/j.rser.2014.02.003.

[11] J. Fisher, N.R. Santen, P. Luckow, F. de Sisternes, T. Levin, A. Botterud, A Guide to Clean Power Plan Modeling Tools - Analytical Approaches for State Plan CO2 Performance Projections, 2016.

[12] M. Ventosa, A. Baíllo, A. Ramos, M. Rivier, Electricity market modeling trends, Energy Pol. 33 (2005) 897–913.

[13] J.F. DeCarolis, K. Hunter, S. Sreepathi, The case for repeatable analysis with energy economy optimization models, Energy Econ. 34 (2012) 1845–1853, https://doi.org/10.1016/j.eneco.2012.07.004.

[14] A.M. Foley, B.P. Ó Gallachóir, J. Hur, R. Baldick, E.J. McKeogh, A strategic review of electricity systems models, Energy 35 (2010) 4522–4530, https://doi.org/10.1016/j.energy.2010.03.057.

[15] J. Timmerman, C. Deckmyn, L. Vandevelde, G. Van Eetvelde, Techno-economic energy models for low carbon business parks, in: Chemical Engineering Transactions, vol. 35, AIDIC - The Italian Association of Chemical Engineering, 2013, pp. 571–576, https://doi.org/10.3303/CET1335095.

[16] K. Mahmud, G.E. Town, A review of computer tools for modeling electric vehicle energy requirements and their impact on power distribution networks, Appl. Energy 172 (2016) 337–359, https://doi.org/10.1016/j.apenergy.2016.03.100.

[17] D. Connolly, H. Lund, B.V. Mathiesen, M. Leahy, A review of computer tools for analysing the integration of renewable energy into various energy systems, Appl. Energy 87 (2010) 1059–1082, https://doi.org/10.1016/j.apenergy.2009.09.026.

[18] L.M.H. Hall, A.R. Buckley, A review of energy systems models in the UK: prevalent usage and categorisation, Appl. Energy 169 (2016) 607–628, https://doi.org/10.1016/j.apenergy.2016.02.044.

[19] S. Jebaraj, S. Iniyan, A review of energy models, Renew. Sustain. Energy Rev. 10 (2006) 281–311, https://doi.org/10.1016/j.rser.2004.09.004.

[20] J. Després, N. Hadjsaid, P. Criqui, I. Noirot, Modelling the impacts of variable renewable sources on the power sector: reconsidering the typology of energy modelling tools, Energy 80 (2015) 486–495, https://doi.org/10.1016/j.energy.2014.12.005.

[21] S.E. Page, What sociologists should know about complexity, Annu. Rev. Sociol. 41 (2015) 21–41, https://doi.org/10.1146/annurev-soc-073014-112230.

[22] K.G. Vamvoudakis, S. Jagannathan, Chapter One - Introduction to Complex Systems and Feedback Control. Control of Complex Systems, Butterworth-Heinemann, 2016, pp. 3–30, https://doi.org/10.1016/B978-0-12-805246-4.00001-X.

[23] A. Klitkou, S. Bolwig, T. Hansen, N. Wessberg, The role of lock-in mechanisms in transition processes: the case of energy for road transport, Environmental Innovation and Societal Transitions 16 (2015) 22–37, https://doi.org/10.1016/j.eist.2015.07.005.

[24] M. Mitchell, Complexity: a Guided Tour, Oxford University Press, Oxford, United Kingdom; New York, 2009.

[25] G.M. Whitesides, R.F. Ismagilov, Complexity in chemistry, Science 284 (1999) 89–92, https://doi.org/10.1126/science.284.5411.89.

[26] W. Brian Arthur, Complexity Economics: A Different Framework for Economic Thought, Complexity and the Economy, 2013.

[27] F. Mazzocchi, Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory, EMBO Rep. 9 (2008) 10–14, https://doi.org/10.1038/sj.embor.7401147.

[28] D. Sherrington, Physics and Complexity, 2009, https://doi.org/10.1098/rsta.2009.0208. ArXiv:09033572 [Cond-Mat].

[29] K.C. Seto, S.J. Davis, R.B. Mitchell, E.C. Stokes, G. Unruh, D. Ürge-Vorsatz, Carbon lock-in: types, causes, and policy implications, Annu. Rev. Environ. Resour. 41 (2016) 425–452, https://doi.org/10.1146/annurev-environ-110615-085934.

[30] D. Möst, W. Fichtner, Einführung zur Energiesystemanalyse (Introduction to energy system analysis), in: Proceedings of the Workshop "Energy System Analysis" of November 27, 2008, at KIT Center Energy, Karlsruhe, Germany, 2009, pp. 11–32.

[31] C. Senkpiel, Akteursverhalten und -strukturen in der Energiesystemanalyse (Actor behavior and structures in energy system analysis), in: Presentation at the Workshop on Complexity in Energy System Analysis, Meeting of the German Research Network on Energy Systems, Berlin, Germany, 2018.

[32] J. Priesmann, L. Nolting, A. Praktiknjo, Are complex energy system models more accurate? An intra-model comparison of power system optimization models, Appl. Energy 255 (2019) 113783, https://doi.org/10.1016/j.apenergy.2019.113783.

[33] L. Nolting, T. Spiegel, M. Reich, M. Adam, A. Praktiknjo, Can energy system modeling benefit from artificial neural networks? Application of two-stage metamodels to reduce computation of security of supply assessments, Comput. Ind. Eng. (2020) 106334.

[34] R. Galvin, How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge? Journal of Building Engineering 1 (2015) 2–12.

[35] J.T. Chi, E.C. Chi, R.G. Baraniuk, k -POD: A method for k -means clustering of missing data, Am. Statistician 70 (2016) 91–99, https://doi.org/10.1080/00031305.2015.1086685.

[36] K. Backhaus, B. Erichson, W. Plinke, R. Weiber, Multivariate Analysemethoden: eine anwendungsorientierte Einführung (Multivariate analysis methods: an application-oriented introduction), in: 15th, Completely Revised Edition, Springer Gabler, Berlin, Germany, 2018.

[37] L. Himmelspach, Vergleich von Strategien zum Clustern von Daten mit fehlenden Werten (Comparison of strategies for clustering data with missing values), in: Proceedings of the 21. GI-Workshop on Foundations of Databases, Rostock-Warnemünde, Mecklenburg-Vorpommern, Germany, June 2-5, 2009, 2009, pp. 129–133.

[38] scikit-learn, K-means Clustering — Scikit-Learn 0.20.2 Documentation, 2017. https://scikit-learn.org/0.19/modules/clustering.html#k-means. accessed February 17, 2019.

[39] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65, https://doi.org/10.1016/0377-0427(87)90125-7.

[40] X. Gao, S. Wu, Hierarchical clustering algorithm for binary data based on cosine similarity, in: 2018 8th International Conference on Logistics, Informatics and Service Sciences (LISS), 2018, pp. 1–6, https://doi.org/10.1109/LISS.2018.8593222.

[41] F. Masulli, S. Rovetta, Clustering high-dimensional data, in: F. Masulli, A. Petrosino, S. Rovetta (Eds.), Clustering High–Dimensional Data, Springer Berlin Heidelberg, 2015, pp. 1–13.

[42] S.-S. Choi, S.-H. Cha, C.C. Tappert, A survey of binary similarity and distance measures, Journal of Systemics, Cybernetics and Informatics 8 (2010) 6.

[43] E. Kaciak, J. Louviere, Multiple correspondence analysis of multiple choice experiment data, J. Market. Res. 27 (1990) 455–465, https://doi.org/10.2307/3172630.

[44] A. Handl, T. Kuhlenkasper, Multivariate Analysemethoden, Springer Berlin Heidelberg, Berlin, Heidelberg, 2017, https://doi.org/10.1007/978-3-662-54754-0.

[45] J.D. Sterman, System dynamics modeling: tools for learning in a complex world, Calif. Manag. Rev. 43 (2001) 8–25, https://doi.org/10.2307/41166098.

[46] M.G. Oladokun, C.O. Aigbavboa, Simulation-Based Analysis of Energy and Carbon Emissions in the Housing Sector, Springer International Publishing, Cham, 2018, https://doi.org/10.1007/978-3-319-75346-1.

[47] B. Dahmen, Kopplung makroskopischer Wirtschaftsverkehrsmodelle mit System Dynamics: Modellierung der Auswirkungen von Umweltzonen auf Fahrzeugflotten (Coupling of macroscopic commercial transport models with system dynamics: Modelling the effects of environmental zones on vehicle fleets), Springer-Verlag, Berlin, Heidelberg, 2019.

[48] Energietechnik: systeme zur konventionellen und erneuerbaren Energieumwandlung. Kompaktwissen für Studium und Beruf (Energy technology: systems for conventional and renewable energy conversion, in: R. Zahoransky (Ed.), Compact Knowledge for Studies and Work), Springer Fachmedien Wiesbaden, Wiesbaden, 2019, https://doi.org/10.1007/978-3-658-21847-8.

[49] L. Nolting, A. Kies, M. Schönegge, M. Robinius, A. Praktiknjo, Locating experts and carving out the state of the art: a systematic review on Industry 4.0 and energy system analysis, Int. J. Energy Res. 43 (2019) 3981–4002, https://doi.org/10.1002/er.4427.

[50] D. Van Hertem, O. Gomis-Bellmunt, J. Liang (Eds.), HVDC Grids: for Offshore and Supergrid of the Future, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2016, https://doi.org/10.1002/9781119115243.

[51] S. Gordon, Supergrid to the rescue, Power Eng. 20 (2006) 30–33, https://doi.org/10.1049/pe:20060505.

[52] BMWi Bundesbericht Energieforschung, Forschungsförderung für die Energiewende 2018 (Federal Report Energy Research 2018 - Research Funding for Energy System Transformation 2018), 2018. https://bit.ly/2UJXAcy. accessed April 06, 2020.

[53] A. Fry, T. Ryley, R. Thring, The influence of knowledge and persuasion on the decision to adopt or reject alternative fuel vehicles, Sustainability 10 (2018) 2997, https://doi.org/10.3390/su10092997.

[54] C.A. Sugar, G.M. James, Finding the number of clusters in a dataset AU - sugar, catherine A, J. Am. Stat. Assoc. 98 (2003) 750–763, https://doi.org/10.1198/016214503000000666.

[55] M. Steinbach, L. Ertöz, V. Kumar, The Challenges of Clustering High Dimensional Data, 2004, https://doi.org/10.1007/978-3-662-08968-2_16.

[56] Peter Zweifel, Aaron Praktiknjo, Georg Erdmann. Energy Economics 1, Springer, Berlin, 2017. ISSN 2192-4333.