Renewable and Sustainable Energy Reviews 158 (2022) 111995

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Modeling flexibility in energy systems — comparison of power sector models based on simplified test cases

Hans Christian Gils [a,b,*], Hedda Gardian [a,b], Martin Kittel [c], Wolf-Peter Schill [c], Alexander Zerrahn [c], Alexander Murmann [d], Jann Launer [e], Alexander Fehler [f], Felix Gaumnitz [f], Jonas van Ouwerkerk [g,h,i], Christian Bußar [g,h,i], Jennifer Mikurda [j], Laura Torralba-Díaz [k,b], Tomke Janßen [l], Christine Krüger [l]

[a] German Aerospace Center (DLR), Institute of Networked Energy Systems, Curiestr. 4, 70563 Stuttgart, Germany
[b] Stuttgart Research Initiative on Integrated Systems Analysis for Energy (STRise), Keplerstraße 7, 70174 Stuttgart, Germany
[c] German Institute for Economic Research (DIW Berlin), Mohrenstraße 58, 10117 Berlin, Germany
[d] Research Center for Energy Economics (FfE), Am Blütenanger 71, 80995 München, Germany
[e] Reiner Lemoine Institute, Rudower Chaussee 12, 12389 Berlin, Germany
[f] Institute for High Voltage Equipment and Grids, Digitalization and Energy Economics (IAEW), RWTH Aachen University, Schinkelstraße 6, 52056 Aachen, Germany
[g] Institute for Power Electronics and Electrical Drives (ISEA), RWTH Aachen University, Jägerstraße 17-19, 52066 Aachen, Germany
[h] Institute for Power Generation and Storage Systems (PGS), E.ON ERC, RWTH Aachen University, Mathieustraße 10, 52074 Aachen, Germany
[i] Jülich Aachen Research Alliance, JARA-Energy, Germany
[j] Chair for Management Science and Energy Economics (EWL), University of Duisburg–Essen, Universitätsstr. 11, 45117 Essen, Germany
[k] Institute of Energy Economics and Rational Energy Use (IER), University of Stuttgart, Heßbrühlstraße 49a, 70565 Stuttgart, Germany
[l] Wuppertal Institute, Döppersberg 19, 42103 Wuppertal, Germany

ARTICLE INFO

Dataset link: https://zenodo.org/record/5802178

Keywords:
Power sector modeling
Model comparison
Sector coupling
Optimization

ABSTRACT

Model-based scenario analyses of future energy systems often come to deviating results and conclusions when different models are used. This may be caused by heterogeneous input data and by inherent differences in model formulations. The representation of technologies for the conversion, storage, use, and transport of energy is usually stylized in comprehensive system models in order to limit the size of the mathematical problem, and may substantially differ between models. This paper presents a systematic comparison of nine power sector models with sector coupling. We analyze the impact of differences in the representation of technologies, optimization approaches, and further model features on model outcomes. The comparison uses fully harmonized input data and highly simplified system configurations to isolate and quantify model-specific effects. We identify structural differences in terms of the optimization approach between the models. Furthermore, we find substantial differences in technology modeling primarily for battery electric vehicles, reservoir hydro power, power transmission, and demand response. These depend largely on the specific focus of the models. In model analyses where these technologies are a relevant factor, it is therefore important to be aware of potential effects of the chosen modeling approach. For the detailed analysis of the effect of individual differences in technology modeling and model features, the chosen approach of highly simplified test cases is suitable, as it allows to isolate the effects of model-specific differences on results. However, it strongly limits the model's degrees of freedom, which reduces its suitability for the evaluation of fundamentally different modeling approaches.

1. Introduction

1.1. Background and motivation

In the European Green Deal, the European Commission has proposed ambitious emission reduction targets for the period from 2021 to 2030 with the aim of achieving climate neutrality by 2050 [1]. To achieve this, the transformation of the energy system towards green technologies has to be accelerated. In the power sector, this requires a switch primarily to variable renewable energy (VRE) technologies such as wind and solar photovoltaics (PV), whose output strongly depends

[*] Corresponding author at: German Aerospace Center (DLR), Institute of Networked Energy Systems, Curiestr. 4, 70563 Stuttgart, Germany.
E-mail address: hans-christian.gils@dlr.de (H.C. Gils).

https://doi.org/10.1016/j.rser.2021.111995
Received 26 March 2021; Received in revised form 31 October 2021; Accepted 9 December 2021
Available online 29 January 2022

List of abbreviations

BEV	battery electric vehicles
BL	base load
CC	controlled charging
CHP	combined heat and power
COP	coefficient of performance
DC	direct current
DR	demand response
EES	electric energy storage
HP	heat pumps
LP	linear programming
LT	long term
MILP	mixed-integer linear programming
PB	peak boiler
PL	peak load
PV	photovoltaics
QP	quadratic programming
ST	short term
TES	thermal energy storage
TPP	thermal power plants
UPGMC	Unweighted Pair-Group Method using Centroids
V2G	vehicle-to-grid
VRE	variable renewable energy

on regional and local weather conditions [2]. As a consequence, the need for system flexibility increases, since power supply and demand have to be balanced in real-time to ensure the security of supply. This flexibility can be provided by different technologies, including controllable power plants, energy storage, transmission grids, or demand-side management [3].

Numerous optimization models have been developed in recent years to provide scientific support in evaluating strategies for the future development of energy supply systems [4]. However, analyses on the future design of the energy system and its operation based on the application of these models usually come to different conclusions [5]. On the one hand, this is driven by different assumptions in the model input data, and on the other hand by differences in model formulations. Models for the analysis of national energy system transformation scenarios usually differ in their spatio-temporal granularity, and technological scope and detail. Limited computational capacities that are still prevailing today, pose a trade-off between these two dimensions [6]. A high spatio-temporal granularity comes at the cost of strong simplifications of the representation of technology properties. These simplifications can differ widely between models. This affects power sector modeling with regard to controllable power plants and combined heat and power (CHP) plants, electric energy storage (EES), transmission grids and demand-side management. The latter includes demand response (DR) of industrial and commercial loads as well as the flexible operation of sector coupling technologies, such as battery electric vehicles (BEV), electric heat pumps (HP) and electrolyzers for the production of hydrogen. The evaluation of different modeling approaches and their impact on results thus requires focused model comparisons that separate the effect of differences in the spatial, temporal, and technological granularity as well as input data used.

1.2. State of research

The literature offers a wide variety of energy system model comparisons (Table 1). These studies can be classified into theoretical comparisons of models (category I), comparisons with a specific technological focus (category II), and comparisons including the harmonized application of different models (category III).

The focus of publications within category I is on comparing a wide range of model functionalities and properties to benchmark and categorize them. This provides energy system modelers and policy makers with a better overview of the existing modeling landscape and supports the selection of a suitable model for a specific research question [14]. Due to the large number of models and the complexity within their implementations a wide variety of studies tries to develop new classification or clustering schemes. Most recent works include Klemm and Vennemann [18] for multi-energy systems, Ridha et al. [15] for complexity comparison, and Prina et al. [16] for bottom-up energy system models. The implications of different modeling approaches on the quality of the results remain largely unclear in those studies.

Category II includes publications that examine specific technical aspects or detailed modeling differences. Their results can help to find the right approaches for future modeling. The comparison, however, is usually based on only one or a few models. To understand the differences between models in depth, a more holistic analysis is required.

In Category III, there are only a few publications with a harmonized, scenario-based comparisons of modeling approaches. Gils et al. [23] performed a systematic comparison with four high resolution power sector models in three scenarios. Siala et al. [24] conducted inter- and intramodel comparisons with five power sector models. Both studies show that even with an unified input data set the results are often not identical. Differences in the implementation of technologies or scenario constraints can lead to a divergent use of flexibility options. However, due to the high complexity of the defined scenarios the causes of the deviations are difficult to investigate.

1.3. Contribution of this paper

Complementing previous literature, this paper is devoted to a systematic, quantitative comparison of optimization and technology modeling approaches in nine models[1]. It is based on a uniform model scope as well as fully harmonized input data. Our work aims at identifying and evaluating the most important differences in the approaches for modeling flexibility in power sector models that include sector coupling options. We systematically contrast optimization and technology modeling approaches, quantify their impact on results and determine pivotal aspects for comparing them across models. To address shortcomings of previous model comparisons, we rely on the analysis of simplified model test cases. To isolate potential differences in results and to analyze their drivers, each test cases is focused on one flexibility option. We model the hourly use of flexibility options over the course of a year, with a focus on supply systems with a high VRE share. While the hourly deployment during one year is endogenously optimized in the model comparison, the available plant capacities are exogenous. The quantitative model comparison is based on standardized indicators representing use patterns of the hourly system operation. Compared to previous work, we include a higher number of models in the comparison. This increases the range of modeling approaches and model features considered, allowing more representative results to be obtained.

The paper is divided into three main parts. Section 2 sets out the methodology of the model comparison. Based on this, Section 3 presents the modeling results and associates differences in results with the model approaches. Finally, Section 4 summarizes and concludes.

[1] The models compared in this paper are modeling frameworks, which allow for modeling a large variety of applications that may differ in terms of spatio-temporal granularity and technological scope. Since this is the much more common term, this text uses a consistent designation as models, not frameworks.

Table 1
Literature overview on the comparison of energy system models.

Reference	Goals and conclusions
Category I: theoretical model comparisons	
[4]	Review of 75 models for the analysis of energy transformation pathways for small-scale to global long-term energy systems. Identifies seven key characteristics pivotal to evaluating VRE integration.
[7]	Comparison of 68 models to assist decision makers in choosing a suitable analysis tool with a focus on integrating VRE into the energy system.
[8]	Comparison of how models address the aspects of temporal and spatial resolution, balancing uncertainties and transparency, growing energy system complexity, and integration of human behavior and social risks. Urges a transformation of models to ensure future applicability.
[9]	Comparison of the representation of EES and transmission networks in long-term electricity models. Concludes that a combination of the advantages of the different model perspectives has not yet taken place.
[10]	Non-comprehensive classification of energy system models in the United Kingdom since 2008. Aims to increase the accessibility of the variety of models both to researchers and policy makers.
[11]	Identification of 67 relevant models that are capable of simulating various aspects with regard to BEV and their integration into power grids.
[12]	Review of 21 expansion planning energy models with a specific focus in policy instruments for VRE integration and decision-support models for energy policy analysis.
[13]	Evaluation of characteristics of national energy system models. Shows that there is a trend to focus on VRE integration. This leads to more flexible approaches with regard to spatial and temporal resolution. Moreover, there is a tendency towards open source.
[14]	Analysis of the ability of energy models to address policy questions. Identifies different terminologies and classification schemes and applies them to 40 selected models.
[15]	Introduction of a clustering approach for energy system models and evaluation of around 150 fact sheets. The main clusters are temporal, spatial, mathematical and modeling content complexity.
[16]	Evaluation of existing classification schemes of bottom-up energy system models. Identifies the concept of resolution as the main indicator. The models in the study show a high resolution in specific fields but lack precision across all fields.
[17]	Identification of seven major challenges in modeling low-carbon energy systems and analyses with a multi-criteria approach, which of 19 models are best suited for addressing those. Finally, it suggests two conceptual modeling suites for bridging the major gaps.
[18]	Identification of models that are suitable to optimize multi-energy systems. Defines a set of characteristics important for modeling them and shows that out of 145 models only few can fulfill the requirements for multi-energy systems optimization.
Category II: specific model comparisons	
[19]	Comparison of three energy models in a case study on the Corvo Island in Portugal. The results show that such models should consider adjustments in their optimization strategies to allow for a better and more cost effective usage of flexible technologies.
[20]	Comparison of linear programming (LP) and mixed-integer linear programming (MILP) formulation for power plants in an hourly-resolved model. It shows that at low VRE shares LP underestimates storage demand, as it neglects technical restrictions that affect operating costs.
[21]	Investigation of the hypothesis that complexity correlates with higher accuracy of results on the basis of 160 modeling configurations. Identifies complexity drivers and model extensions that contribute to significant result accuracy.
[22]	Analysis of the applicability of expansion planning models. Evaluates advantages and disadvantages of the three defined model categories optimization model, equilibrium models and alternative models without an optimal VRE integration.
Category III: comparisons including the harmonized application of different models	
[23]	Evaluation of three sector-coupled power systems for Germany in 2050 using four different models. The paper highlights the importance of harmonized input data and the need for simplified test cases for gaining detailed insight into the impact of model differences.
[24]	Evaluation of the impact of model type, planning horizon, temporal and spatial resolution by comparing five power sector models with harmonized input data and characteristics. Concludes that harmonization is crucial for understanding deviations in results.

Table 2
Overview of investigated test cases and contributing model versions. A cross (X) stands for the participation of the models in the test cases. A dot (•) means that the models offer the possibility to consider the analyzed flexibility options, but this was not applied in the test cases investigated here.

Test case label	Analyzed flexibility option	DIETER	E2M2	GENESYS-2	ISAaR	JMM	MarS	oemof	REMix	RESTORE
TPP(PL/LP)	peak load (PL) power plants LP	X	X	X	X	X	X	•	X	X
TPP(BL/LP)	base load (BL) power plants LP	X	X	X	X	X	X	•	X	
TPP(BL/MILP)	BL power plants MILP		X			X	X	•	•	
ResHydro	Reservoir hydro power	X	X	X	•	X	X	X	X	•
EES(ST)	short term (ST) EES	X	X	X	•	X	X	•	X	X
EES(LT)	long term (LT) EES	X	X	X	•		X	•	X	X
EES(ST+LT)	ST and LT EES	X	X	X	•		•	•	X	X
PowGrid	Power transmission	X	X	X	X	X	X	X	X	X
DR	Demand response	X	•			•	X	•	X	X
HP+TES	Electric HP with thermal storage[a]	X	•		X	X		X	X	•
BEV(CC)	BEV with controlled charging[a]	X	•		•	X	X	•	X	X
BEV(V2G)	BEV with bidirectional charging[a]	X	•		•	X	X	X	X	X
H2+Stor	Hydrogen electrolyzers with storage[a]	•			X	•	X	•	X	X
CHP(BP)	Backpressure (BP) CHP		X		X	X	X	X	X	
CHP(Ex)	Extraction (Ex) CHP		X		X	X	X	X	X	
CHP(BP)+PB	Backpressure CHP with peak boiler		•		X	X		X	X	
CHP(Ex)+PB	Extraction CHP with peak boiler		•		X	X		X	X	
FlexHeatNetw	Flexible heating network		•		X	•		X	X	

[a] Peak load power plants are also available.

2. Materials and methods

This section describes the framework of the model comparison. First, the procedure for conducting the model comparison and the data used are described in Sections 2.1 and 2.2, respectively. Second, the models involved are introduced in Section 2.3, and their main differences are outlined in Section 2.4. Third, the indicators used for the comparison are characterized in Section 2.5.

2.1. Set-up of the model comparison

The model test cases represent a highly simplified system consisting of electricity demand, VRE power generation from wind and PV, and one, in exceptional cases two, flexibility options. In addition, in the sector coupling test cases, heat demand, hydrogen demand, and BEV charging electricity are considered. To match demand and generation, the models can use the respective flexibility option as well as VRE curtailment and uncontrolled load shedding. However, the latter is associated with very high costs. In total, we separately consider 18 flexibility options in individual test cases (Table 2). They focus on different types of thermal power plants (TPP) and EES, electricity transmission grids, DR in industry and commerce, and various flexible sector coupling technologies. The latter includes electric air-to-water building HP with thermal energy storage (TES), BEV with controlled charging (CC) and vehicle-to-grid (V2G), decentralized hydrogen electrolysis with tank storage, and CHP. CHP is analyzed for backpressure plants and extraction condensation plants, in each case separately for stand-alone plants and plants combined with a peak boiler (PB). In addition, extraction condensation CHP is also analyzed as part of a flexible heat network with PB, electric boiler, HP and TES. Since these only provide demand-side flexibility but not a power supply option, the test cases on HP, BEV, and hydrogen electrolysis include peak load power plants in addition to renewable power generation.

2.2. Input data of the model comparison

The model comparison is based on a uniform input data set that is used in all models. It defines the exogenous installed generation capacities considered in each case as well as their techno-economic parameters, the energy demands, and various time series. The time series indicate the hourly course of the demand for electricity, heat and hydrogen, the electricity generation of wind power and PV, the inflow to reservoir hydro power plants, and the flexibility of BEV and DR. As input data requirements differ across models, some models may not use all technology-specific data points.

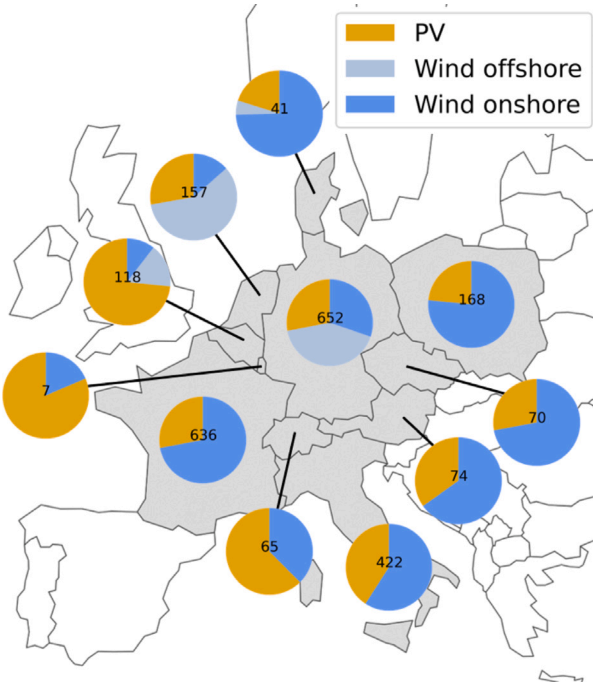

Fig. 1. Model regions considered in the comparison (gray) with VRE generation shares depicted in pie charts and the total annual VRE generation in TWh in each diagram. These shares and values apply to the test cases without additional electricity demand due to sector coupling. Even though measured demand profiles and historical weather years are used for the countries shown, the test cases are highly stylized. For example, currently existing capacities of hydro power plants and thermal power plants are not considered, and the transmission grid is only considered in one test case.

The regional scope of the test cases includes 11 regions. For parts of the input data used, such as electricity demand or VRE generation profiles, these regions correspond to different countries in Central Europe (Fig. 1). However, the modeled system is not meant to be a real representation of their energy system, as we neglect currently existing power generation and storage capacities. Instead, we consider stylized plant capacities for the modeled technologies. Also, the existing power grid is only considered in the test case focused on power transmission. Thus, these are exemplary model regions with different amounts and temporal profiles of demand and VRE generation.

The assumed wind and PV capacities are identical across all models. However, there are differences between the test cases, as the additional

Table 3
Overview of programming procedures and optimization approaches of the contributing models as they are used in the comparison.

	DIETER	E2M2	GENESYS-2	ISAaR	JMM	MarS	oemof	REMix	RESTORE
Modeling language	GAMS	GAMS	C++	MATLAB, PostgreSQL	GAMS	Fortran	Python	GAMS	MATLAB
Problem formulation[a]	LP	LP, MILP	population-based heuristic	LP	LP, MILP	MILP, DP, Lagrange	LP	LP	QP
Foresight in hours	8760	8760	1	8760	24/36	8760	8760	8760	144
Objective	min. costs	min. costs	min. costs	min. costs	min. costs	min. costs	min. costs	min. costs	min. residual load
Objective function	OPEX	OPEX	OPEX	OPEX	OPEX	OPEX, Lagrange multipliers	OPEX	OPEX	residual load balancing
Documentation	[28,29]	[30,31]	[32,33]	[34–36]	[37]	[38]	[39–41]	[42–44]	[45,46]

[a] LP — linear programming, MILP — mixed-integer linear programming, DP — dynamic programming, QP — quadratic programming.

electricity demand of sector coupling must be accompanied by higher VRE capacities to realize uniform supply shares. To minimize the number of different input data sets, the electricity demands of HP, BEV and electrolysers are assumed to be identical. The corresponding hydrogen and heat demands were then calculated using the efficiency of the electrolyzers and the coefficient of performance (COP) of the HP, respectively. The heat demand to be met by CHP or flexible heat networks is assumed to be identical to that of HP.

The VRE capacities corresponding to the two input data sets of electricity demand are calculated in separate upstream optimization runs with the REMix model. We make an exogenous assumption of a theoretical VRE supply share of 80%, which could only be realized if curtailment and losses were completely avoided. As no power grid is considered when determining these capacities, the supply share of 80% applies to each of the model regions. Since the regions have different VRE potentials in terms of installed capacity and hourly electricity generation, different optimal combinations of PV, wind onshore and wind offshore result (Fig. 1). The techno-economic parameters are assumed identical for all regions.

The assumed capacity of flexibility options is identical to the maximum residual load to be covered in the case of TPP, CHP, reservoir hydro power, and EES. Transmission capacities are assumed to slightly increase compared to today, as they are expected by [25] for the year 2030. We do not differentiate capacities by flow direction as this is not possible in all models. Instead, the larger of the two values is used in each case. For DR, potentials for load shifting and controlled load shedding in industry and commerce are set according to [26]. In the test cases with one of the sector coupling technologies, the capacities of the peak load power plants are adapted to the – in these cases higher – residual peak load.

To enable automated processing by the models, the model input data is provided in a uniform template. This template is available together with the used model input data at [27].

2.3. Contributing models

The models involved in the comparison are hourly resolved multi-node power sector optimization models with representation of different flexibility options. However, there are numerous differences in terms of programming procedures and objective function (Table 3). While in previous applications, the models were used to analyze systems with different geographic scope and spatial detail, here we aim for a fully harmonized application. Not all models are used in every test case (Table 2). This is partly due to the scope of the respective model, but also partly due to the scope of the project, which strives for a modeling effort that is as uniform as possible.

As Table 3 shows, the majority of contributing models minimize total system costs, which here only include the operational costs (OPEX), under perfect foresight. However, three models have fundamentally different optimization approaches as specified in the following.

Heuristic dispatch model approach. In GENESYS-2, a dispatch model provides a fixed technology dispatch order for every time step. A distinction is made between two different system states: either there is a VRE surplus (negative residual load) or there is a VRE shortfall (positive residual load). In case of a negative residual load, the surplus initially is balanced across regions if possible. Subsequently, short-term EES are charged until they reach full charge capacity, then charging of these units is possible with an additional cross-regional balancing. The same procedure is then applied to long-term EES. Remaining surplus is curtailed. In case of a positive residual load, there is an equivalent procedure. The model balances the shortfall across regions if possible. Then, it discharges storage starting with short-term storage. In a last step, the model operates TPP to cover the remaining residual load.

Rolling horizon approach. JMM uses a rolling planning horizon to optimize the yearly dispatch. The year is divided into shorter periods to reduce the size of the optimization problem, and, therefore, the resulting overall computation time. Additionally, this approach offers the opportunity to consider information updates like in case of renewable forecasts. In JMM, every 12 hours (h) a new optimization period starts with a length of 24 or 36 h.

RESTORE also uses the rolling horizon approach. The optimization period is set to 72 h, the step size to 36 h. Furthermore, an aggregated foresight horizon enables a longer-term forecast: hours that lie after the actual optimization period are aggregated and appended (for most of the cases considered here: 72 h, aggregated into 6 clusters). Generally, models with reduced foresight are limited in optimal storage use over longer periods of time. To consider the long-term use of storage, a filling level must be specified for the end of each optimization period. In RESTORE this is realized through a separate, upstream module. Here, a single year-round optimization with perfect foresight with reduced temporal resolution is done. These results are then set as constraints for the detailed optimization with rolling horizon, ensuring seasonal effects are considered. Otherwise, instead of retaining the stored energy for usage at a later point in time, a complete discharge of the stored energy would be incentivized.

Quadratic residual load minimization approach. In contrast to all other contributing models, RESTORE minimizes the positive residual load and not the system cost. In doing so, the model maximizes the VRE use and minimizes the required back-up capacity without considering economic restrictions of flexibility options. Beyond that, it uses a quadratic programming (QP) approach instead of a linear one to avoid load peaks and reduce gradients.

2.4. Technology modeling differences

Beyond the optimization approaches, there are a number of differences that affect technology modeling as outlined in the following. The overview in Table 4 focuses on the differences that are essential for the subsequent comparison of the results. For technologies not listed there, no relevant differences between the participating models were identified.

Thermal power plants. A wide range of different approaches emerges when considering load change constraints and costs for TPP. In the case without integer variables, there are some models that do not foresee load change constraints and costs at all (GENESYS-2, oemof). In others, the load change incurs additional costs that scale linearly with the hourly load change (DIETER, E2M2, ISAaR, REMix). In JMM, these costs only apply for started up capacities. In addition to load change costs, further costs incur in ISAaR when power plants leave a certain capacity range and fuel consumption is higher at partial load. Similar to ISAaR, also in E2M2 and JMM a higher fuel consumption at partial load as well as for starting-up capacities is taken into account. In contrast to this, in MarS load change constraints are considered, however no load change costs are applied. In E2M2 load change and start-up constraints and costs are considered only in test cases modeling base load power plants.

In addition, in the case of a MILP formulation a minimal power feed-in has to be maintained during operation (E2M2, JMM, MarS). Further restrictions apply considering minimum up and down times of TPP. Moreover, E2M2 and MarS also include explicit ramping restrictions. In JMM, the power plant restrictions are not only applied for the specific MILP test case, but are also used in a modified way for the LP test cases. Similar as in JMM these restrictions are also considered in E2M2 with a LP formulation but only in test cases modeling base load power plants.

The unavailability of TPP can be modeled either by a continuous power reduction based on a given availability rate or by a stochastic approach. Within the stochastic approach, which is exclusively considered in MarS, discrete units are randomly drawn and made unavailable reducing the overall available generation capacity. Thus, unlike the other models, the hourly values are not an exogenous assumption. On average over the entire year, the plant availability corresponds to the constant values.

Hydro power plants. Hydro power plants are characterized by a set of reservoirs subject to natural inflows, which are interconnected by turbines and pumps. In most models, a simplified implementation using an aggregated approach is applied. Interconnected storage reservoirs, inflows, turbines, and pumps are combined in one common unit. Further differences in the modeling approaches exist regarding the consideration of pumps and natural inflows. The DIETER model version used here does not consider pumping. In GENESYS-2 direct inflows to the reservoirs cannot be implemented. In all other models both direct inflows as well as pumps are modeled simultaneously. In contrast to the aggregated models, in MarS a more detailed model of the sequential interconnection of reservoirs is implemented, considering the water masses, which are circulated between individual reservoirs. This structure allows that water masses that flow through multiple reservoirs and turbines could generate electricity multiple times.

Electric energy storage. The basic representation of EES is very similar for all participating models. The most relevant difference is the consideration of minimum initial and final storage levels in some models (Table 4). These storage boundaries have not been harmonized. Apart from the fact that this is not possible in all models due to different model requirements, the aim here is to identify the differences in model results arise from these different model formulations.

Power transmission. The consideration of power transmission lines differs primarily in the modeling approach. While REMix uses a direct current (DC) load flow approach [47], all other models employ a simplified transport model [48]. The difference is limited to the model formulation; REMix also does not represent a detailed network topology. Another model difference concerns the consideration of transmission losses, which are accounted for in all models except MarS and RESTORE.

Demand response. The main differences in the model representation concerns the basic approach of modeling DR either as a storage technology with additional time constraints (RESTORE), or as a load shifting process that explicitly constrains loads shifted in specific hours (DIETER, MarS and REMix). Further, the models differ with respect to limitations of the usage frequency, maximum shifting duration, intervals between load interventions, and time-variable availability of the potential or energy losses. In the RESTORE model, DR is implemented as a virtual energy storage with time-variable boundary conditions for power and energy according to the methodology described in [49]. Differing from this, MarS uses a generic load shifting model. Loads can be shifted within a defined time window, and the hourly shift potential can be specified. In contrast, DR availability is assumed to be time-invariant in DIETER. Here, load increases or decreases have to be balanced within a symmetrical maximal shift duration either prior to or after an intervention. Furthermore, DIETER includes a regeneration time for each shifted energy unit, which has to elapse before the next load shift [50]. In contrast to the other models, REMix uses fixed shift durations [51]. This implies that when the load changes, it is already determined when the compensation takes place. To limit the size of the model, not all possible values up to the maximum shift duration are usually considered. Furthermore, intervention durations, frequencies and regeneration times between interventions are limited by approximated energy quantities of the load shift, which are calculated from mean values of the potentials. Temporal load shifting is considered in all models that contribute to the DR test case, controlled load shedding only in DIETER and REMix. Losses are considered in all models except MarS. DR costs are incurred in all models except RESTORE.

Battery electric vehicles. For the sake of comparison, BEV are represented as one 'swarm' aggregate of vehicle load and storage in all contributing models. The implementation largely coincides, with subtle differences regarding technical and economic restrictions. BEV entail a flexible charging of the vehicle's batteries (BEV-CC), and an additional flexible generator that reconverts the battery's energy back into the electricity grid (BEV-V2G). BEV that are not being connected to the grid are assumed to be driving on the road. For most models, the time-variant driving profile induces a variable electricity demand supplied by the batteries. In contrast, the JMM model assumes that, before disconnecting, individual BEV batteries are fully charged and, after driving, vehicles return to the grid with a pre-defined storage level. All models restrict the battery capacity by time-variant minimum and maximum load levels, aggregated over the sum of BEV. The maximum level is defined by the number of vehicles that are connected to the grid and their specific battery capacity. In case of JMM and RESTORE models, the swarm battery does not need to retain a minimum level as safety margin, but can be fully discharged. Further differences are attributed to the costs of (dis)charging. The REMix model penalizes deviations from an exogenous profile, which refers to uncontrolled charging. All other contributing models do not impose such penalties. Variable costs for charging and/or discharging energy apply in DIETER, JMM, and oemof. REMix also considers discharging costs.

Combined heat and power. As with TPP, the modeling of CHP plants differs in the degrees of constraints and costs of ramping as well as the unit availability. A complementary feature is that some models (ISAaR, oemof) offer the possibility of excess CHP electricity generation, which results in increased flexibility. Furthermore, in one model (MarS), the interaction on the heat side is not explicitly modeled. Instead, it is translated into must-run electric generation, resulting in increased flexibility compared to explicit modeling.

Electric heat pumps. Some of the models (DIETER, oemof, REMix) include a temperature-dependent COP, using ambient temperature time series, whereas others (ISAaR, JMM) assume a constant COP throughout the year. In DIETER and JMM, all heat produced has to go through the attached TES. In contrast, ISAaR, oemof and REMix feature a bypass. The availability of a storage bypass is also relevant in the test case considering hydrogen electrolysis (Table 4).

Table 4
Overview of technology modeling differences and features relevant for the result comparison.

Technology	DIETER	E2M2	GENESYS-2	ISAaR	JMM	MarS	oemof	REMix	RESTORE
Power plant ramping (where applicable including CHP)	simple load change costs	operational restrictions, part load efficiencies	no flexibility constraints or costs	simple load change costs, linearized part load behavior	operational restrictions, part load efficiencies	efficiency depending on operation point	no flexibility constraints or costs	simple load change costs	power plants not explicitly modeled
Power plant unavailability	constant	constant	constant	constant	constant	stochastic model of outages	constant	constant	
Reservoir hydro power	aggregated, w/ inflow, w/o pumping	aggregated, w/ pumping and inflow	aggregated, w/o inflow, w/ pumping	aggregated, w/ pumping and inflow	aggregated, w/ pumping and inflow, generation based on internal water value calculation	Hydraulic networks consisting of interconnected reservoirs, turbines and pumps	aggregated, w/ pumping and inflow	aggregated, w/ pumping and inflow, [52]	aggregated, w/ inflow, w/o pumping
Start and end storage levels	start: 50 %, end: 50 %	optimized, equal	start: 0 %, end: optimized	start: 0 %, end: optimized	start: 50 %, end: optimized	start: 50 %, end: 50 %	optimized, equal	optimized, equal	start: 0 %, end: optimized
Demand response	maximum shifting times, regeneration time, time-invariant shifting potentials [50]					defined shifting time frames, in which the energy has to be compensated. No limits in frequency.		time-variant potential, fixed shifting and intervention times, limits in frequency [51]	implemented as storage with time-variable boundaries [49]
Heat pumps	temperature-dependent COP [53]			constant COP	constant COP		temperature-dependent COP	temperature-dependent COP	temperature-dependent COP
Storage bypass (thermal and/or hydrogen)	w/o bypass			w/ bypass	w/ bypass (CHP), w/o bypass (HP)		w/ bypass	w/ bypass [26]	w/o bypass, thermal storage implicitly modeled as thermal load shiftability
Power transmission	NTC-based w/ losses	NTC-based w/ losses	NTC-based w/ losses	NTC-based w/ losses	NTC-based w/ losses	NTC-based w/o losses	NTC-based w/ losses	DC load flow w/ losses	NTC-based w/o losses
Battery electric vehicles	with minimum battery level, CC and V2G costs scale with total charged or discharged energy				w/o minimum battery level, vehicle must be fully charged before disconnecting, no CC but V2G costs	with minimum battery level, no CC and V2G costs	with minimum battery level, no CC but V2G costs	with minimum battery level, CC cost scale with deviations from exogenous charging profile	without minimum battery level, no costs for CC and V2G

2.5. Output indicators

The evaluation of the model comparison focuses on the use of the available flexibility options. In a broader sense, this also includes curtailment of VRE generation and uncontrolled load shedding. The latter is implemented in the models as a slack variable to ensure the balance of power, heat, and hydrogen to keep the mathematical problem solvable. These two indicators are complemented by the system costs and flexibility usage. Depending on the flexibility option, this is represented by electricity, heat or hydrogen generation, storage utilization, storage and grid losses, load shifting, and transmitted electricity.

Besides scalar indicators, we analyze hourly use profiles of plant operation. In particular, the use of flexibility options, but also of VRE curtailment and uncontrolled load shedding (corresponding to uncovered load), is compared for selected times of the year. This allows the observation of deviating plant usage behavior.

To enable an automated evaluation of the results, the output variables of all models are transferred into a standardized data format, which is then read by the evaluation scripts.

We use a cluster analysis tool to identify systematic result deviations. The tool applies a hierarchical cluster algorithm (Unweighted Pair-Group Method using Centroids (UPGMC)), which allocates model results on region and indicator level to clusters based on the similarity of results. The latter is determined by a pair-wise distance matrix indicating the Euclidean distance between all result values. First, the pair of most similar models, i.e., those with lowest Euclidean distance between their result values, are joined together to one cluster. Subsequently, the UPGMC algorithm updates the distance matrix, including the distance between the newly formed cluster and all other result values using the cluster centroid. This is the arithmetic mean of all results grouped in one cluster. In the next cluster step, again the most similar pair of result values or clusters are joined together, with a subsequent update of the distance matrix. Again, the procedure repeats and terminates at a predetermined distance threshold. The resulting cluster structure of models supports the identification of structural consistencies of models grouped together in one cluster, and differences between models grouped in different clusters. Models in singleton clusters with only one cluster member are considered outliers, indicating either erroneous model parameterization that can be improved, or structural discrepancy to all other models.

3. Results and discussion

The analysis of the four key indicators, curtailment, uncovered load, system costs, and flexibility usage, shows that the range of results varies greatly depending on the test case. This implies that the identified model differences (Tables 3 and 4) have very diverse impacts, depending on the flexibility option considered.

The absolute values of the indicators are closely linked to the characteristics of the flexibility option examined. Thus, large amounts of uncovered load occur primarily when technologies cannot provide controllable electricity or can do so only to a limited extent (DR, PowGrid, EES). In turn, high VRE curtailment occurs when other constraints limit the flexibility of electricity supply (CHP(BP), CHP(Ex)). Both uncovered load and VRE curtailment are also reflected in the flexibility usage. In addition, uncovered load is a very strong driver of system costs. Not only in absolute terms, but also in the observed ranges of differences in results, there are corresponding dependencies between the indicators.

Depending on the possibility and extent of a provision of positive and/or negative balancing energy, the considered flexibility options have different effects on the system. From this follows that the values of flexibility usage shown in Fig. 2(d) are not always directly comparable. This applies in particular to cases in which several interrelated flexibility options are available to the system (FlexHeatNetw, BEV(V2G)).

We analyze the differences in hourly plant deployment for selected test cases by calculating the correlation of the hourly values between the model pairs and averaging them over all time steps of the year (Fig. 3). Particularly high correlations are found for the dispatch of peak load power plants, charging of BEV, and operation of HP, whereas the spread of model results is much larger for long-term storage, BEV grid feed-in, and DR. With regard to long-term storage, however, it should be noted that the differences result primarily from a constant

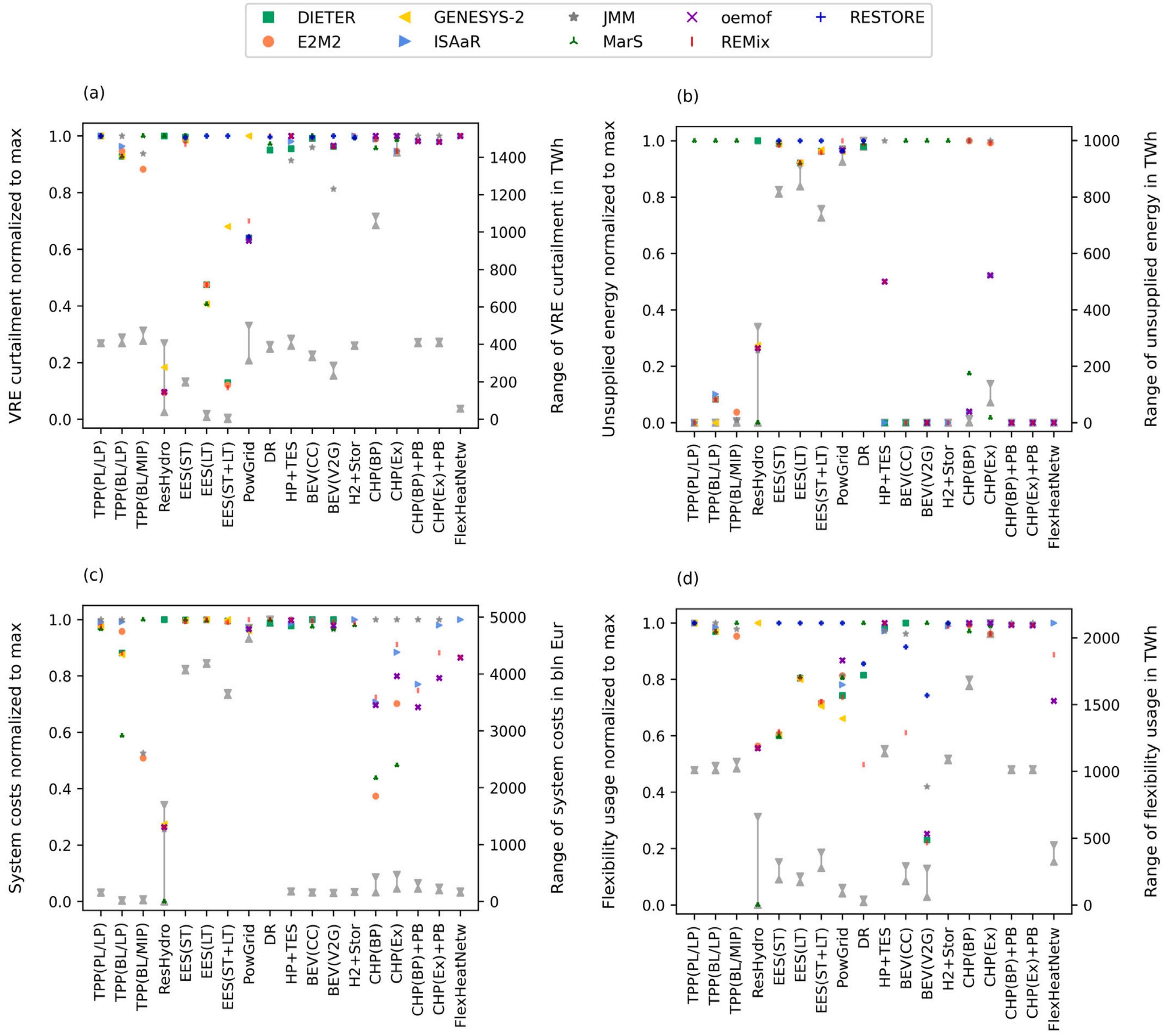

Fig. 2. Overview of the model results for the temporarily and spatially aggregated main indicators. The colored symbols show the individual model results, normalized with the maximum value (left axis). The lines indicate the range of absolute values (right axis). The subfigures illustrate annual VRE curtailment (a), annual sum of uncontrolled load shedding (b), total system costs (c), and flexibility usage (d). The latter represents the power production of power plants for the test cases TPP, ResHydro, CHP, HP+TES, H2+Stor; storage output for test cases EES, and FlexHeatNetw; power transmission for PowGrid; load shifting/controlled load shedding for DR, and BEV(CC); and grid feed-in for the test case BEV(V2G). Uncontrolled load shedding corresponds to an uncovered load and includes electricity, heat and hydrogen, depending on the test case. Table 2 details the test cases.

offset, since the annual sum of electricity supply shows a high agreement in most models (Fig. 2(d)). The opposite effect is particularly evident in the case of the HP, where a significantly different operating behavior occurs despite relatively small differences in the annual sums. Fig. 4 gives an example of the differences in the hourly operation of selected flexibility options.

Results on the flexibility usage suggest a clustering of our 18 test cases into three categories. In the cases analyzing TPP, CHP (both with and without peak boiler) and hydrogen electrolysis (H2+Stor), there are only minor differences in the range of a few percent or TWh across the models, mostly caused by individual model features. This usually involves the use of a few additional constraints or model parameters while maintaining the same basic approach to technology modeling. Substantial deviations can be observed in the test cases with EES, power transmission (PowGrid), building HP, and the flexible heating network (FlexHeatNetw). They predominantly result from the different optimization approaches. Results diverge to the largest extent in the test cases focusing on DR, BEV, and reservoir hydro power (ResHydro). They are driven by fundamentally different approaches of technology modeling.

We elaborate on the results following the above-mentioned categorization logic (driven by differences in technology modeling, optimization approaches and model features). Additionally, we associate result deviations to the model differences identified in Section 2.

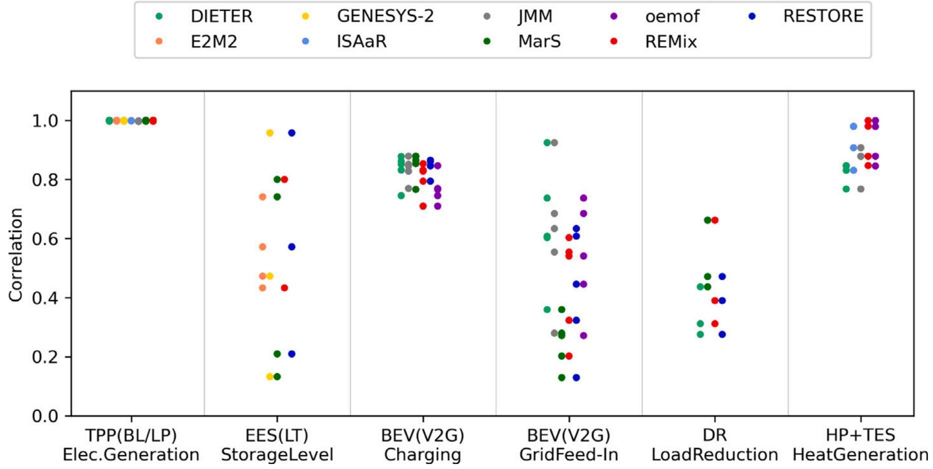

Fig. 3. Correlation of model time series for selected test cases and parameters, where the scatter points for each model represent the correlation to the other model's plant operation over 8760 time steps.

3.1. Impact of different technology modeling approaches

Power transmission. When using a DC load flow approach (REMix), not all lines can be utilized simultaneously according to their nominal capacity. Thus, balancing of positive and negative peaks of the residual load requires more line capacity compared to the simplified transport model (all other models). As a result, the amount of electricity transmitted decreases by about 10% in the corresponding test case, with the VRE curtailment increasing by approximately the same value (Fig. 2). Furthermore, the uncovered load and, thus, system costs increase by about 3%. In order to separate the interaction of the different network representations with model differences in other technologies, power transport between regions is not possible in all other test cases.

Demand response. As a consequence of the different modeling approaches (Section 2.4), there are strong deviations in the annual energy quantities of load shifting and controlled shedding (Fig. 2(d)), but also in the hourly operations (Fig. 3 and Fig. 4(e)). In the absence of any other flexibility option, DR costs should not drive results. Thus, differences in usage result from modeling. Considering fixed shift durations, maximum usage durations, and/or frequency constraints results in a 3–4 times lower and more time-variable usage (DIETER, REMix). In contrast, modeling DR as a time-constrained storage technology allows for usage durations of multiple hours and increased shifted energy amounts (RESTORE). Not considering regeneration periods and daily maxima increases the frequency of DR use by up to a factor of 10 (MarS). Greater amounts of shifted load cause a stronger reduction of VRE curtailment and uncovered load (Fig. 2). Another finding is that, in the case of a DR potential that varies strongly over time, a consideration of daily energy quantity maxima on the basis of average values substantially reduces load shifting (REMix) compared to the other models. Consideration of losses, on the other hand, is not essential for the results at the values examined.

Battery electric vehicles. Higher flexibility, i.e. fewer limitations in terms of economic or technical restrictions, promotes BEV charging behavior that is beneficial to VRE power generation. The different constraints on minimum storage levels and cost assumptions for CC and V2G (Section 2.4) lead to a widely differing flexibility usage (Fig. 2). Imposing costs on the deviation from an exogenously specified charging profile (REMix) results in a different behavior especially in times of low VRE availability (Fig. 4(c) and (d)), as compared to approaches that do not incentivize a certain charging profile (all other models). In contrast, usage of V2G is mostly driven by differences in the optimization approaches (Section 3.2).

Reservoir hydro power. The consideration of a detailed cascading model (MarS) enables a higher electricity generation by hydro turbines since the water masses can be used several times for power generation when flowing through multiple reservoirs and turbines. This lead to a 10% increased generation in MarS compared to models with the aggregated approach (E2M2, JMM, oemof, REMix). Therefore, uncovered load and pumping can be completely avoided within the cascading model, which is not the case for models applying the aggregated approach (Fig. 2). These differences in the results are explicitly based on the actual implementation, since both approaches are characterized by the same overall storage and conversion capacity. The strong upward outliers in VRE curtailment, uncovered load, and system costs (Fig. 2) results from the disregard of pumping (DIETER).

3.2. Impact of different optimization approaches

The usage of a fixed order of dispatch without temporal foresight (GENESYS-2) leads to substantially different results only in the test cases with EES. Part of the explanation is that the usage of storage technologies not only depends on the maximum available power output but also on the storage level. The storage level is in turn influenced by the charging and discharging strategy and short planning horizon. This strategy is markedly different and less efficient compared to models using perfect foresight. It drives an increase in curtailment and uncovered load, reflected also in the operational behavior of EES. The maximum deviation can be observed for long-term storage with the curtailment rising about 50% compared to LP models.

Perfect foresight over a whole year should, in principle, allow for a more efficient operation of the optimized system than models with rolling planning horizon (JMM, RESTORE) — provided that restrictions of the modeled technologies extend over several, individually optimized time steps. For a duration of the individual optimization steps of 72 h plus 72 h aggregated foresight (RESTORE), this here only applies to long-term storage, whose deployment is optimized in a separate modular procedure (Section 2.3), which strongly reduces the impact of the rolling planning approach.

The shorter time horizon of 24 to 36 h (JMM) is not considered in the test case with long-term storage, but in those evaluating battery storage, TES, and BEVs. However, remuneration of the storage filling level at the end of every optimization period in the objective function yields similar results in comparison to models with perfect foresight. This is related to the fact that, in the simplified test cases, VRE surplus can either be stored or curtailed. In more complex systems, other flexibility options or power plants with detailed operating restrictions offer additional applications for excess electricity and, therefore, probably

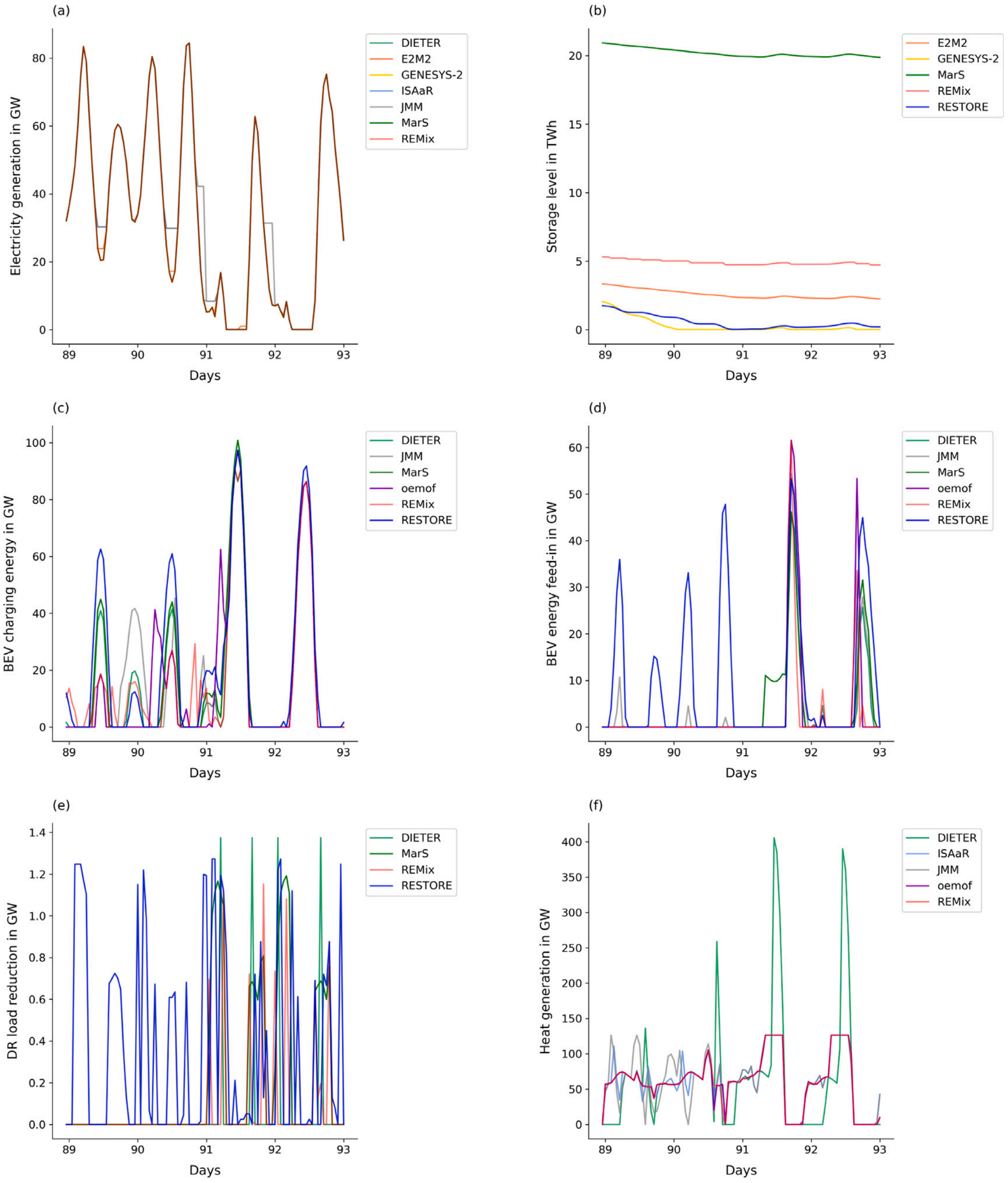

Fig. 4. Time series of selected indicators using the example of a spring period and the model region representing Germany. Shown are the operation of base load TPP (a), the filling of long-term EES (b), charging (c) and grid feed-in (d) of BEVs, load reduction for one of the DR technologies (e), and building HP operation (f). The differences in technology dispatch during the four days shown result from the VRE power generation. While the first two days are relatively windless, the following days are characterized by surplus situations at midday due to a higher wind power feed-in. This results in a different flexibility requirement, which leads to a higher or lower effect of the model differences depending on the technology considered.

more substantial model result differences. Only in the test case BEV (V2G), reduced curtailment can be observed in JMM, which is expected to be mainly caused by different modeling restrictions for BEV and not by the rolling planning horizon.

The application of quadratic objective function substantially affects the technology dispatch by smoothing out peak loads with priority. We observe an additional usage of storage, power transmission, BEV flexibility, and DR in the corresponding test cases. This is accompanied in most cases by relatively high VRE curtailment, and in the case of long-term EES also by larger amounts of unsupplied energy. As a result, there are also differences in the timing of V2G use (Fig. 4).

3.3. Impact of model features

Besides the more fundamental differences in optimization and technology modeling approaches, a variety of smaller model features have an impact on the results.

Grid losses. Neglecting losses in power transmission leads to a reduction of VRE curtailment and uncovered load, which however has only minor effects on the system costs in the test case considered here.

Initial and end storage levels. Depending on their implementation (Table 4), initial and end storage levels may have a pronounced impact on the results for long-term EES. Usually, long-term storage filling levels follow a seasonal pattern, driven by the availability of the dominating

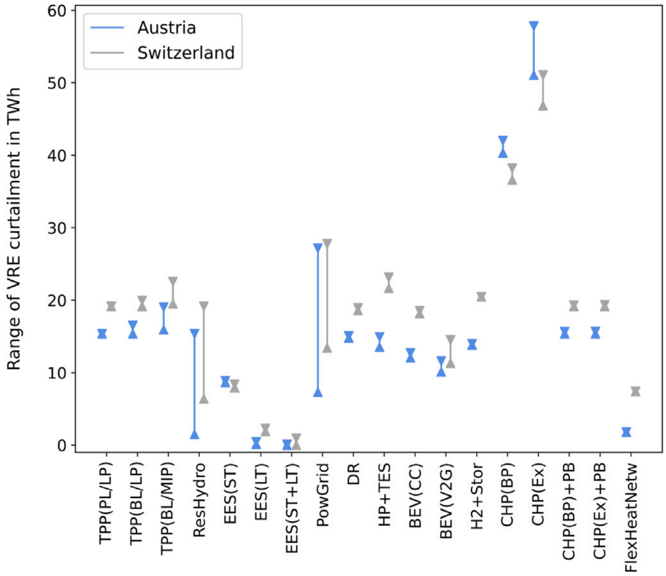

Fig. 5. Comparison of the range of model results for the indicator of VRE curtailment for the two model regions Austria and Switzerland. In Austria, 65% of the VRE power generation originate from wind onshore and 35% from PV, while the shares for Switzerland are almost the opposite (38% and 62%). The annual demand to be covered is about 15% higher in Austria (Fig. 1).

energy source. Fig. 4 shows the impact of different initial and end storage levels for long-term storage. The definition of initial storage levels (RESTORE, GENESYS-2, MarS) leads to a divergent mode of operation. Allowing a lower storage level at the end of the year than at the beginning can also have an impact on storage operation: This can be observed, for example, in the test case BEV (CC), where JMM uses part of the energy initially present in the storage to reduce electricity production.

Power plant outages. The consideration of stochastic outages (MarS) causes larger amounts of uncovered load if they coincide with a high residual peak load. This is relevant for the test cases of TPP and CHP and all others in which peak load power plants are taken into account (Fig. 2). In the case of CHP, however, this is partially superimposed by the disregard of a CHP power curtailment, and in the case of base load TPP by the constraints and costs of power plant ramping.

Constraints and costs of power plant ramping. Despite the numerous differences in the approaches to modeling constraints and costs of TPP ramping, the spread of results is relatively small. This is due to the limited number of degrees of freedom for the optimization if only one flexibility option is available. The modeling differences are most noticeable in the case of base load TPP, which are characterized by a low technical flexibility (Fig. 4). Additional constraints and costs of ramping can reduce the amount of electricity provided by up to 5%, resulting in an increase of curtailment, uncovered load and system costs (Fig. 2). Gas-fired power plants are also considered in the test cases of HP, BEV and electrolyzers. Since there are few other model differences, at least in the case of HP and electrolyzers, the consideration of TPP ramping is a relevant driver of the differences in results there.

In the test cases analyzing CHP, ramping constraints and costs also cause a deviating operational behavior. However, the amounts of electricity provided are almost identical. In the test case with a flexible heating network, the CHP ramping has a substantial impact on the interaction of CHP, HP, and TES. Strong temporal changes in the residual load are preferably compensated by adjusting the HP input in case of additional CHP ramping restrictions, which favors a more intense usage of the TES (ISAaR).

Combined heat and power. The possibility of excess CHP electricity (ISAaR, oemof) allows for the provision of additional heat in cases where there are no other heat generators. This is reflected in substantially lower values for uncovered load, here for heat, in the CHP(BP) and CHP (Ex) test cases (Fig. 2). However, since the case of an isolated CHP plant is very contrived, this should not affect more realistic model applications, as the test cases with alternative heat sources show. Furthermore, increased CHP plant ramping costs result in very slightly lower CHP usage in cases with peak boiler (ISAaR).

Time-variant COP. Models that represent air-to-water building HP with a temperature-dependent and time-variant COP (DIETER, oemof, REMix) have about 8% higher electricity consumption than those using a yearly-averaged scalar value for COP. This is due to the fact that time-variant COP are lower in winter due to colder ambient temperature, which is the time of the year with the highest heating demand. The modeling of a time-variable COP is therefore of high relevance to avoid an underestimation of the generation capacity required in cold winter hours. The hourly operation pattern shows a clear concentration of HP operation to the hours with a higher COP especially on winter days, increasing the use of the TES (Fig. 4).

Storage bypass. As expected, disallowing the option of a bypass for thermal or hydrogen storage leads to considerably higher values for charging and discharging. Besides a difference in reporting, this can have an impact on the results when variable operational costs or charge/discharge losses apply.

3.4. Interaction of model features and data

When analyzing the results, systematic effects of the interaction between the input data or the selected test case and the model differences become apparent. In general, the observed differences in the results are clearly driven by the chosen design of the simplified test cases and its limited technology portfolio. Thus, while the simplified design allows for a fairly good association of model and result differences, it also has a limiting effect on the possibility of generalizing the results. For example, the differences between linear minimization of costs and quadratic minimization of residual load turn out to be quite small when only one flexibility option is available. Similarly, the impact of a fixed dispatch order and a rolling horizon approach is rather limited in our test cases with few degrees of freedom. Without considering the interaction of different flexibility options, there is a fairly intensive use of the available technology in each case. Thus, the quantitative differences in results can only be transferred to more comprehensive scenarios to a limited extent.

Furthermore, the restriction to one flexibility option can lead to non-unique optima due to situations where multiple options with the same cost or residual load level are available. In our test cases this applies, in particular, to the question in which hour and for which technology VRE curtailment occurs during a period of consecutive hours with renewable surplus. While this does not affect the annual totals of flexibility deployment and other indicators, it may result in different hourly patterns. Identifying this effect in a reliable way proved to be non-trivial in the analysis. It may be avoided by considering random noise costs of VRE curtailment.

Complementary to these overarching effects, some model differences may affect some model region stronger than others. For instance, this concerns the effect of stochastic power plant outages. It is found to be most effective in regions with smaller total installed generation capacities, where the failure of individual units has a relatively larger impact. While there are large relative differences in the uncovered load, absolute figures are very small (Fig. 2). Furthermore, there are systematic dependencies between the VRE supply structure and the range of model results for curtailment (Fig. 5). The relative deviations between the models tend to be larger when technologies are particularly suitable for balancing the respective dominant VRE technology. This presumably results from the fact that the model differences become more pronounced with a more frequent technology deployment. In the case of wind-dominated supply, the results in the test cases with long-term EES, reservoir hydro power, base load TPP, hydrogen electrolysis, CHP with boiler, controlled BEV charging, and HP spread more widely. In the case of power generation predominantly from PV, on the other hand, greater dispersion is observed for short-term EES and V2G. No clear trend is observed for CHP without boiler and power transmission. These observations result from comparing several pairs of countries with comparable absolute demand, each with contrasting and clearly pronounced dominance of one VRE technology (Fig. 6). In contrast, a comparison of the ranges of results for VRE curtailment for model regions with similar supply structures but widely varying amounts of demand reveals no systematic trends.

3.5. Recommendations for future model comparisons

Based on the experience of previous model comparisons [23], a relatively large amount of time was invested in a theoretical comparison of the models in preparation for the modeling work. Thereby, the required input data of each model was gathered and model overviews were generated. These helped to understand the different models and facilitated the analysis.

In addition, the exact design of the data interfaces and the naming of the parameters under consideration is important to establish a common understanding. This includes input as well as output parameters. Prior standardization can reduce the need for repetitive model runs caused by different interpretations of model input parameters or errors in the transfer from the input database. Especially the usage of harmonized data formats has proven to be beneficial.

To ensure the plausibility of the considered test cases, it is advisable to first test them in one model before rolling them out to all models. Furthermore, the extensive automation of model parameterization and evaluation via automated interfaces and scripts for data processing proved to be very helpful in making the large number of models and repeated calculations manageable.

It was also useful to develop a routine that automatically gathered all relevant results, and created standardized figures. In addition, it was beneficial that the analysis could be performed in a decentralized way by all participants. In this context, it seems advisable to include central model input variables, such as installed plant capacities, in the evaluation. In doing so, parameterization errors can be identified more quickly.

For the comparison of results, a combined analysis of annual aggregated results as well the time series is advantageous allowing investigations of differences in the operation of technologies. To quickly identify similarities in results, a cluster analysis is useful. However, for some technologies, such as BEV, the analysis of differences in results over time also proved to be difficult due to systemic interactions of various model differences.

4. Conclusions

To quantify the understanding of the effect of fundamental but also small-scale modeling decisions on the results of temporally and spatially resolved power system models, our work was dedicated to the detailed analysis of nine models and their application in fully harmonized but highly stylized test cases. In doing so, a significant effort had to be made for the complete harmonization of the models. This harmonization has contributed significantly to the mutual validation of the models.

The initial comparison of the general model characteristics showed that each of them is characterized by certain properties regarding optimization approach and technology modeling. However, a detailed comparison revealed that many of the models do not differ from each other in the way individual technologies are modeled. Pronounced differences in technology modeling were identified primarily for DR, BEV, reservoir hydro power, and power transmission. In model analyses where these technologies are a relevant factor, it is therefore important to be aware of potential effects of the chosen modeling approach. More specifically, the comparison of DR modeling suggests that the use of explicit shift durations and usage constraints is particularly important when considering real-world processes. Simplified approaches are well suited for evaluating the potential of aggregate load flexibility. When modeling the flexibility of BEV, the consideration of costs of controlled charging as well as vehicle-to-grid proves to be particularly relevant. This should be considered accordingly in analyses with a focus on electric mobility. Furthermore, a detailed consideration of the cascades of storage hydro power plants is recommended when a dedicated analysis of individual plants or systems is the focus. In contrast, the aggregated approach is sufficient for an approximate assessment of the role of hydro power in integrated future energy systems. Because it considers the interaction of power flow across all lines connected to a node, the DC load flow approach is more suited when the focus is on analyzing the use of existing grid connections, especially in the evaluation of critical supply situations. In contrast, the transport model approach, which overestimates real flows, is sufficient for aggregated system planning.

In addition to these more pronounced differences in technology modeling, a wide range of minor differences in model features was identified in the analysis, such as the consideration of grid losses or a temperature dependent COP of HP. However, these have only a limited impact on model outcomes. It can be concluded that the model features often capture complementary constraints or effects of technology dispatch, and thus allow for a more detailed analysis. They should be considered in focused analyses of individual technologies.

With respect to technology modeling, our analysis indicates that a detailed representation is most important when a flexibility option is particularly suitable for balancing the usual generation characteristics of the VRE technology prevalent in the system under consideration.

The comparison with previous model comparisons reveals that the use of fully harmonized input data as well as the consideration of reduced test cases allows for a dedicated analysis of specific model differences. However, it also becomes apparent that, despite the very simplified test cases, some technologies have multiple and, in some cases, interlinked degrees of freedom. Overlapping effects are a major challenge for interpreting model results and cannot always be separated and quantified. This especially holds true for the interpretation of hourly profiles of plant operation characterized by the interaction of several model differences. Nonetheless, our analysis indicates that when comparing results of temporally resolved models, not only aggregated annual values but also time series have to be considered. For example, despite similar annual aggregates of technology use, completely different usage patterns can occur, but the opposite can also be observed. Additionally, it must be examined individually whether a deviating operation can also result from non-unique solutions, or such solutions are to be prevented by suitable methods.

With regard to the fundamentally different modeling approaches (QP, fixed dispatch order, rolling horizon), our approach proves to be of limited suitability for model comparison. Since only few degrees of freedom are available to the models in the simplified test cases, these approaches yield only minor deviations in the results. Rather, the test cases do not allow the models to manifest their respective strengths. This is for quadratic optimization the analysis of a maximum residual load smoothing, for the fixed dispatch order the possibility of an integral analysis of hourly resolved transformation pathways and for the rolling horizon the consideration of a limited temporal foresight corresponding to reality. Using simplified test cases may therefore not be suitable to determine the effects of the differences of these approaches, since these might rather appear in more complex settings.

In complementary future work, it would be desirable to investigate to what extent the identified relationships between model properties and differences in results can be transferred to more complex scenarios. This includes the further exploration of overlapping effects. Finally, an extension to scenarios with endogenous capacity expansion could provide complementary insights.

CRediT authorship contribution statement

Hans Christian Gils: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Hedda Gardian:** Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Martin Kittel:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Wolf-Peter Schill:** Methodology, Investigation, Writing – review & editing, Funding acquisition. **Alexander Zerrahn:** Methodology, Software, Validation, Formal analysis, Investigation. **Alexander Murmann:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **Jann Launer:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Alexander Fehler:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Felix Gaumnitz:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Jonas van Ouwerkerk:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Christian Bußar:** Writing – review & editing, Funding acquisition. **Jennifer Mikurda:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Laura Torralba-Díaz:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Tomke Janßen:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Christine Krüger:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft , Writing – review & editing, Funding acquisition.

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.

Data availability

The input data and the data template used are available on https://zenodo.org/record/5802178.

Acknowledgments

The research for this paper was performed within the FlexMex project supported by the German Federal Ministry of Economic Affairs and Energy under grant number 03ET4077. The authors thank Georgios Savvidis, Felix Böing, Kai Hufendiek, Benjamin Fleischer, Christoph Weber, Berit Müller, Mascha Richter, Frank Merten, Dirk Uwe Sauer, Niklas van Bracht, Arne Pöstges, and Caroline Podewski for contributing to the project conceptualization and funding acquisition.

Appendix A

See Fig. 6.

References

[1] European Commission. A European green deal: Striving to be the first climate-neutral continent. 2021, https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en, last accessed on 22/02/2021.

[2] López Prol Javier, Schill Wolf-Peter. The economics of variable renewable energy and electricity storage. Annu Rev Resour Econ 2021;13(1):443–67. http://dx.doi.org/10.1146/annurev-resource-101620-081246.

[3] Zöphel Christoph, Schreiber Steffi, Mueller Theresa, Moest Dominik. Which flexibility options facilitate the integration of intermittent renewable energy sources in electricity systems? Curr. Sustain./Renew Energy Rep 2018;5:37–44. http://dx.doi.org/10.1007/s40518-018-0092-x.

[4] Ringkjøb Hans-Kristian, Haugan Peter M, Solbrekke Ida Marie. A review of modelling tools for energy and electricity systems with large shares of variable renewables. Renew Sustain Energy Rev 2018;96:440–59. http://dx.doi.org/10.1016/j.rser.2018.08.002.

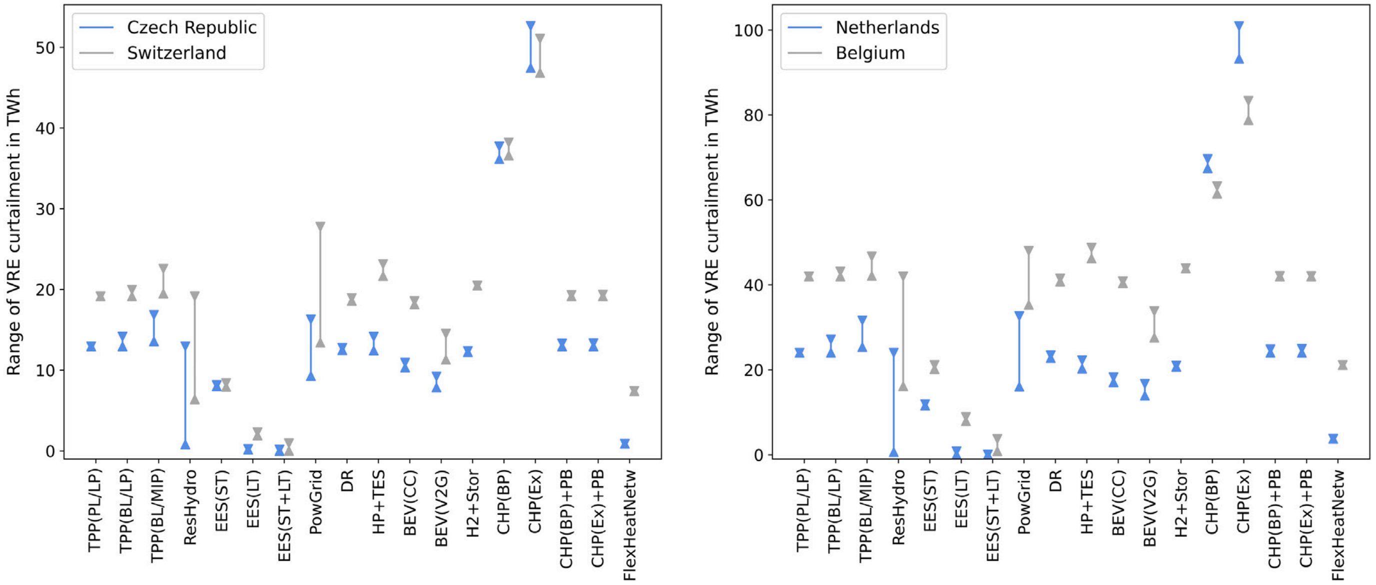

Fig. 6. Comparison of the range of model results for the indicator of VRE curtailment for the two model regions Czech Republic and Switzerland (left) as well as Belgium and the Netherlands (right). PV supply shares are 73% in Belgium and 28% in the Netherlands, wind shares are 27% and 72%, respectively.

[5] Naegler Tobias, Sutardhio Claudia, Weidlich Anke, Pregger Thomas. Exploring long-term strategies for the german energy transition - A review of multi-sector energy scenarios. Renew Sustain Energy Transition 2021;100010. http://dx.doi.org/10.1016/j.rset.2021.100010.

[6] Cao Karl-Kiên, von Krbek Kai, Wetzel Manuel, Cebulla Felix, Schreck Sebastian. Classification and evaluation of concepts for improving the performance of applied energy system optimization models. Energies 2019;12(24). http://dx.doi.org/10.3390/en12244656.

[7] Connolly David, Lund Henrik, Mathiesen Brian Vad, Leahy Martin. A review of computer tools for analysing the integration of renewable energy into various energy systems. Appl Energy 2010;87:1059–82. http://dx.doi.org/10.1016/j.apenergy.2009.09.026.

[8] Pfenninger Stefan, Hawkes Adam, Keirstead James. Energy systems modeling for twenty-first century energy challenges. Renew Sustain Energy Rev 2014;33:74–86. http://dx.doi.org/10.1016/j.rser.2014.02.003.

[9] Després Jacques, Hadjsaid Nouredine, Criqui Patrick, Noirot Isabelle. Modelling the impacts of variable renewable sources on the power sector: Reconsidering the typology of energy modelling tools. Energy 2015;80:486–95. http://dx.doi.org/10.1016/j.energy.2014.12.005.

[10] Hall Lisa MH, Buckley Alastair R. A review of energy systems models in the UK: Prevalent usage and categorisation. Appl Energy 2016;169:607–28. http://dx.doi.org/10.1016/j.apenergy.2016.02.044.

[11] Mahmud Khizir, Town Graham E. A review of computer tools for modeling electric vehicle energy requirements and their impact on power distribution networks. Appl Energy 2016;172:337–59. http://dx.doi.org/10.1016/j.apenergy.2016.03.100.

[12] Gacitua Leonardo, Gallegos Pablo, Henriquez-Auba Rodrigo, Lorca Álvaro, Negrete-Pincetic Matías, Olivares Daniel, et al. A comprehensive review on expansion planning: Models and tools for energy policy analysis. Renew Sustain Energy Rev 2018;98:346–60. http://dx.doi.org/10.1016/j.rser.2018.08.043.

[13] Lopion Peter, Markewitz Peter, Robinius Martin, Stolten Detlef. A review of current challenges and trends in energy systems modeling. Renew Sustain Energy Rev 2018;96:156–66. http://dx.doi.org/10.1016/j.rser.2018.07.045.

[14] Savvidis Georgios, Siala Kais, Weissbart Christoph, Schmidt Lukas, Borggrefe Frieder, Kumar Subhash, et al. The gap between energy policy challenges and model capabilities. Energy Policy 2019;125:503–20. http://dx.doi.org/10.1016/j.enpol.2018.10.033.

[15] Ridha Elias, Nolting Lars, Praktiknjo Aaron. Complexity profiles: A large-scale review of energy system models in terms of complexity. Energy Strategy Rev 2020;30. http://dx.doi.org/10.1016/j.esr.2020.100515.

[16] Prina Matteo Giacomo, Manzolini Giampaolo, Moser David, Nastasi Benedetto, Sparber Wolfram. Classification and challenges of bottom-up energy system models - a review. Renew Sustain Energy Rev 2020;129. http://dx.doi.org/10.1016/j.rser.2020.109917.

[17] Fattahi Amirhossein, Sijm Jos, Faaij André. A systemic approach to analyze integrated energy system modeling tools: A review of national models. Renew Sustain Energy Rev 2020;133. http://dx.doi.org/10.1016/j.rser.2020.110195.

[18] Klemm Christian, Vennemann Peter. Modeling and optimization of multi-energy systems in mixed-use districts: A review of existing methods and approaches. Renew Sustain Energy Rev 2021;135(110206). http://dx.doi.org/10.1016/j.rser.2020.110206.

[19] Neves Diana, Pina Andre, Silva Carlos A. Demand response modeling: A comparison between tools. Appl Energy 2015;146:288–97. http://dx.doi.org/10.1016/j.apenergy.2015.02.057.

[20] Cebulla Felix, Fichter Tobias. Merit order or unit-commitment: How does thermal power plant modeling affect storage demand in energy system models? Renew Energy 2017;105:117–32. http://dx.doi.org/10.1016/j.renene.2016.12.043.

[21] Priesmann Jan, Nolting Lars, Praktiknjo Aaron. Are complex energy system models more accurate? An intra-model comparison of power system optimization models. Appl Energy 2019;255:113783. http://dx.doi.org/10.1016/j.apenergy.2019.113783.

[22] Dagoumas Athanasios S, Koltsaklis Nikolaos E. Review of models for integrating renewable energy in the generation expansion planning. Appl Energy 2019;242:1573–87. http://dx.doi.org/10.1016/j.apenergy.2019.03.194.

[23] Gils Hans Christian, Pregger Thomas, Flachsbarth Franziska, Jentsch Mareike, Dierstein Constantin. Comparison of spatially and temporally resolved energy system models with a focus on Germany's future power supply. Appl Energy 2019;255. http://dx.doi.org/10.1016/j.apenergy.2019.113889.

[24] Siala Kais, Mier Mathias, Schmidt Lukas, Torralba-Díaz Laura, Sheykkha Siamak, Savvidis Georgios. Which model features matter? An experimental approach to evaluate power market modeling choices. 2020, arXiv:2010.16142.

[25] ENTSO-E: TYNDP - market modelling data. 2016, URL https://www.entsoe.eu/Documents/TYNDP%20documents/TYNDP%202016/rgips/TYNDP2016%20market%20modelling%20data.xlsx, last accessed on 20/01/2022.

[26] Gils Hans Christian. Balancing of Intermittent Renewable Power Generation by Demand Response and Thermal Energy Storage. (Ph.D. thesis), University of Stuttgart; 2015, http://dx.doi.org/10.18419/opus-6888.

[27] Hedda Gardian, Gils Hans Christian, Kittel Martin, Murmann Alexander, Launer Jann, Gaumnitz Felix, Fehler Alexander, van Ouwerkerk Jonas, Mikurda Jennifer, Torralba-Díaz Laura, Krüger Christine, Janßen Tomke, Zerrahn Alexander. Model input and output data of the FlexMex model comparison. 2021, http://dx.doi.org/10.5281/zenodo.5802178.

[28] Zerrahn Alexander, Schill Wolf-Peter. Long-run power storage requirements for high shares of renewables: review and a new model. Renew Sustain Energy Rev 2017;79:1518–34. http://dx.doi.org/10.1016/j.rser.2016.11.098.

[29] Gaete-Morales Carlos, Kittel Martin, Roth Alexander, Schill Wolf-Peter. DIETERpy: A python framework for the dispatch and investment evaluation tool with endogenous renewables. SoftwareX 2021;15:100784. http://dx.doi.org/10.1016/j.softx.2021.100784.

[30] Sun Ninghong. Modellgestützte Untersuchung des Elektrizitätsmarktes: Kraftwerkseinsatzplanung und -Investitionen. (Ph.D. thesis), Institute of Energy Economics and Rational Energy Use, University of Stuttgart; 2013, http://dx.doi.org/10.18419/opus-2159.

[31] Torralba-Díaz Laura, Schimeczek Christoph, Reeg Matthias, Savvidis Georgios, Deissenroth-Uhrig Marc, Guthoff Felix, et al. Identification of the efficiency gap by coupling a fundamental electricity market model and an agent-based simulation model. Energies 2020;13(15). http://dx.doi.org/10.3390/en13153920.

[32] Bußar Christian. Untersuchung optimaler Transformationspfade bis 2050 für die erfolgreiche Umsetzung einer nachhaltigen Reduzierung der Kohlendioxidemissionen im Bereich der Stromerzeugung. (Ph.D. thesis), RWTH Aachen University; 2019, http://dx.doi.org/10.18154/RWTH-2019-09975.

[33] Siemonsmeier Marius, Bracht Niklas, Bußar Christian. Transformation des Energiesystems mit steigendem Anteil Erneuerbarer Energien mit Netz- und Speicherausbau unter einer gesamteuropäischen Perspektive; Thema: Entwicklung eines Simulationsprogramms und Untersuchung von Energieversorgungsszenarien. RWTH Aachen University; 2018, http://dx.doi.org/10.2314/KXP:1687958572.

[34] Böing Felix, Murmann Alexander, Pellinger Christoph, Kigle Stephan. ISAaR - Integrated simulation model for unit dispatch and expansion with regionalization. Tech. rep., FfE; 2019, URL https://www.ffe.de/en/isaar.

[35] Pellinger Christoph. Mehrwert Funktionaler Energiespeicher aus System- und Akteurssicht. (Ph.D. thesis), Technical University of Munich; 2016, URL https://mediatum.ub.tum.de/1303981.

[36] Böing Felix. Cross-sector assessment of CO2 abatement measures and their impact on the transmission grid. (Ph.D. thesis), Technical University of Munich; 2020, URL https://mediatum.ub.tum.de/1539536.

[37] Meibom Peter, Barth Rüdiger, Hasche Bernhard, Brand Heike, Weber Christoph, O'Malley Mark. Stochastic optimization model to study the operational impacts of high wind penetrations in Ireland. IEEE Trans Power Syst 2011;26(3):1367–79. http://dx.doi.org/10.1109/TPWRS.2010.2070848.

[38] Drees Tim. Simulation des europäischen Binnenmarktes für Strom und Regelleistung bei hohem Anteil erneuerbarer Energien. (Ph.D. thesis), RWTH Aachen University; 2015, URL https://publications.rwth-aachen.de/record/658687.

[39] Hilpert Simon, Kaldemeyer Cord, Krien Uwe, Günther S, Wingenbach Clemens, Pleßmann Guido. The open energy modelling framework (oemof) - A new approach to facilitate open science in energy system modelling. Energy Strategy Rev 2018;22:16–25. http://dx.doi.org/10.1016/j.esr.2018.07.001.

[40] Krien Uwe, Schönfeldt Patrik, Launer Jann, Hilpert Simon, Kaldemeyer Cord, Pleßmann Guido. Oemof.solph—A model generator for linear and mixed-integer linear optimisation of energy systems. Softw Impacts 2020;6:100028. http://dx.doi.org/10.1016/j.simpa.2020.100028.

[41] Welcome to oemof's documentation! — oemof.solph 0.4.2.dev0 documentation, URL https://oemof-solph.readthedocs.io/en/latest/, last accessed on 20/01/2022.

[42] Gils Hans Christian, Scholz Yvonne, Pregger Thomas, de Tena Diego Luca, Heide Dominik. Integrated modelling of variable renewable energy-based power supply in europe. Energy 2017;123:173–88. http://dx.doi.org/10.1016/j.energy.2017.01.115.

[43] Gils Hans Christian, Simon Sonja. Carbon neutral archipelago - 100% renewable energy supply for the canary islands. Appl Energy 2017;188:342–55. http://dx.doi.org/10.1016/j.apenergy.2016.12.023.

[44] Gils Hans Christian, Gardian Hedda, Schmugge Jens. Interaction of hydrogen infrastructures with other sector coupling options towards a zero-emission energy system in germany. Renewable Energy 2021;180:140–56. http://dx.doi.org/10.1016/j.renene.2021.08.016.

[45] Krüger Christine, Buddeke Mathis, Merten Frank, Nebel Arjuna. Modelling the interdependencies of storage, DSM and grid-extension for europe. In: 12th International conference on the european energy market (EEM) : 19-22 May 2015, Lisbon. New York, NY: Inst. of Electrical and Electronics Engineers; 2015, http://dx.doi.org/10.1109/EEM.2015.7216669.

[46] Buddeke Mathis, Krüger Christine, Merten Frank. Modellbeschreibung: Einsatzmodell für Flexibilitätsoptionen im europäischen Stromsystem. Tech. rep., Wuppertal Institute for Climate, Environment, Energy; 2016, URL https://wupperinst.org/fa/redaktion/downloads/projects/Restore2050_AP7_Modell.pdf, last accessed on 20/01/2022.

[47] Cao Karl-Kiên, Metzdorf Johannes, Birbalta Sinan. Incorporating power transmission bottlenecks into aggregated energy system models. Sustainability 2018;10(6). http://dx.doi.org/10.3390/su10061916.

[48] Cao Karl-Kiên, Pregger Thomas, Haas Jannik, Lens Hendrik. To prevent or promote grid expansion? Analyzing the future role of power transmission in the European energy system. Front Energy Res 2021;8:371. http://dx.doi.org/10.3389/fenrg.2020.541495.

[49] Kleinhans David. Towards a systematic characterization of the potential of demand side management. 2014, arXiv:1401.4121.

[50] Zerrahn Alexander, Schill Wolf-Peter. On the representation of demand-side management in power system models. Energy 2015;84:840–5. http://dx.doi.org/10.1016/j.energy.2015.03.037.

[51] Gils Hans Christian. Economic potential for future demand response in Germany-modeling approach and case study. Appl Energy 2016;162:401–15. http://dx.doi.org/10.1016/j.apenergy.2015.10.083.

[52] Scholz Yvonne. Renewable energy based electricity supply at low costs : development of the REMix model and application for Europe. (Ph.D. thesis), University of Stuttgart; 2012, http://dx.doi.org/10.18419/opus-2015.

[53] Schill Wolf-Peter, Zerrahn Alexander. Flexible electricity use for heating in markets with renewable energy. Appl Energy 2020;266. http://dx.doi.org/10.1016/j.apenergy.2020.114571.