Renewable and Sustainable Energy Reviews 159 (2022) 112177

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Model-related outcome differences in power system models with sector coupling—Quantification and drivers

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

[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

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.5802178

Keywords:
Power sector modeling
Model comparison
Sector coupling
Optimization

ABSTRACT

This paper presents the results of a multi-model comparison to determine outcome deviations resulting from differences in power system models. We apply eight temporally and spatially resolved models to 16 stylized test cases. These test cases differ in their renewable energy supply share, technology scope, and optimization scope. We focus on technologies for balancing the variability of power generation, such as controllable power plants, energy storage, power transmission, and flexibility related to sector coupling. We use harmonized input data in all models to separate model-related from data-related outcome deviations. We find that our approach allows for isolating and quantifying model-related outcome deviations and robust effects concerning system operation and investment decisions. Furthermore, we can attribute these deviations to the identified model differences. Our results show that trends in the use of individual flexibility options are robust across most models. Moreover, our analysis reveals that differences in the general modeling approach and the modeling of specific technologies lead to comparatively small deviations. In contrast, a heterogeneous model scope can cause substantially larger deviations. Due to a large number of models and scenarios, our analysis can provide important information on which investment and operation decisions are robust to the model choice, and which modeling approaches have an exceptionally high impact on results. Our findings may guide both modelers and decision-makers in properly evaluating the results of similarly designed power system models.

1. Introduction

1.1. Background and motivation

Optimizing system models are among the standard tools used for energy systems analysis. Such models are often applied for investigating future energy supply systems. Considering power supply, the integration of fluctuating power generation from variable renewable energy (VRE) through flexibility options such as storage, grids, and controllable power plants is the focus of many models and their application [1]. This is increasingly complemented by analyses of the flexibility that can be tapped when implementing so-called sector coupling [2]. This essentially refers to the direct and indirect use of electricity in other areas of the energy system to reduce greenhouse gas emissions there. In this context, the partially flexible use of electricity for charging battery electric vehicles (BEVs) [3], for heat generation

* 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.2022.112177
Received 29 August 2021; Received in revised form 20 January 2022; Accepted 20 January 2022
Available online 17 February 2022
1364-0321/

List of abbreviations

BEV	Battery electric vehicle
CHP	Combined heat and power
CO_2	Carbon dioxide
DC	Direct current
DH	District heating
E2P	Energy to power
HP	Heat pump
IAM	Integrated assessment model
PV	Photovoltaics
TES	Thermal energy storage
VRE	Variable renewable energy

in heat pumps (HPs) and electric boilers [4], and the electrolytic generation of hydrogen [5] are of great importance.

However, scenarios based on the application of such models often come to different conclusions regarding future technology use. Differences in model outcomes can result from any step in the modeling chain [6], e.g. from different assumptions regarding the development of demand, costs and technology parameters (data-related differences), but also from different modeling approaches (model-related differences). The fundamental cause of divergent model outcomes is the necessity to abstract complex systems as part of the modeling, which can be realized in different ways. The mathematical formulation of the model plays a role here, as do the scope and detail of the spatial, temporal and technological model dimensions. A comprehensive understanding of the effects of particular modeling approaches is thus a prerequisite for the correct interpretation of the scenarios and model results. Structured model comparisons are a helpful tool to gain this understanding. To quantify model-related deviations in outcomes, data-related differences must be kept as small as possible by using harmonized input data [7].

1.2. State of research

The literature offers a range of publications on comparisons of energy or power system models. However, these are mostly limited to a theoretical comparison of the methodology used, the model scope, or the model properties. The most recent of these papers focus, for example, on the consideration of policy instruments in the models [8], the technological focus [9], the ability to address policy-relevant research questions [10], the comparison of model resolutions [11], the ability to analyze renewable energy systems [12], or the optimization of multi-energy systems [13]. There is no application of the compared models in any of these publications.

In contrast, Sugiyama et al. [14] present a comparative application of spatially and temporally aggregated energy system models to transformation scenarios for Japan. However, an input data harmonization, an analysis of model differences and a comparison of model properties were not conducted. Similarly, a set of eight models differing widely in their temporal and spatial detail was applied to a scenario analysis for the North-American energy system in [15]. Their comparison is also not based on fully harmonized input data, and the differences in results are not related to model properties. North America is also the assessment area of another comparison considering 17 models and 13 scenarios [16]. Again, there is no harmonization of the input data and no analysis of model-related outcome deviations. The model comparison of Giarola et al. [17] is also devoted to North America, but examines future energy storage expansion in particular. Since the four models used have numerous differences in scope and input data, there are extensive result deviations, which can only be partially attributed to model characteristics.

A coordinated application of four models to three scenarios of a future German power system, including flexible sector coupling and regional resolution, was carried out in Gils et al. [18]. In Siala et al. [19], a systematic comparison of the effects of different model types, planning horizons, spatial and temporal resolutions was performed applying five models to different power system scenarios for Germany. Both works rely on harmonized model input data, but do not provide a systematic analysis of model-related outcome deviations considering multiple scenarios. The deployment pathways of VRE technologies in the United States of America were the focus of another comparison of three energy planning models [20]. Despite the use of harmonized input data, large ranges of plant expansion result there, which is attributed to different technology modeling of VRE. Ruhnau et al. [21] compare five power system models with partially harmonized model configurations and inputs. They identify the representation of combined heat and power (CHP) as one key driver of result differences at high VRE shares. In another comparison of four power system models with harmonized input data, Misconel et al. [22] evaluate result differences caused by diverging modeling approaches. Both papers do not consider sector coupling and only analyze individual scenarios for the year 2030. Gils et al. [7] applied nine models with fully harmonized input data to highly simplified test cases to explore the effect of model differences in detail. On this basis, the impact of differences in the model formulation can be well understood for individual technologies.

There is a wide range of publications in which multiple models are applied to transformation scenarios in integrated assessment modeling. For example, [23] focuses in particular on VRE integration modeling, [24] on energy technology cost assumptions, [25] on carbon price impacts, and [26] on national contributions for the achievement of the Paris agreement. All these works have in common that different models are applied to the respective scenarios considered, but without harmonization of input data and detailed exploration of differences in results. Methods for the harmonization of input data of integrated assessment models (IAMs) are addressed in some recent works. Krey et al. [27] conduct a review on techno-economic parameters in IAMs, and encounters significant differences. They also identify numerical differences in technology modeling as another possible cause of differences. However, data harmonization as well as model application are not performed. Giarola et al. [28] address how such a harmonization could be implemented and which challenges would have to be overcome. They show that the differences in IAM results can potentially be reduced with the application of the developed framework for data harmonization.

In summary, previous comparisons of energy system models are predominantly of theoretical nature and do not include a harmonized application of the models. Where an application of the models has been implemented, differences in results were not systematically explored and attributed to model properties. In addition, complete harmonization of input data and model configurations is done only in very few cases. Similarly, model-related differences in results are not captured for individual technologies. Furthermore, considering flexible sector coupling plays no or only a minor role in earlier comparisons of power system models.

1.3. Contribution of this paper

Complementing the existing literature, our paper systematically assesses model-related differences in power system models, considering stylized future energy scenarios. In particular, we focus on technologies for balancing fluctuating power generation from VRE, referred to as flexibility options. These include electricity storage, transmission grids, and flexible sector coupling. Given this focus, our comparison includes a portfolio of eight power system models[1] optimizing one year of

[1] Models compared in this paper are frameworks that enable the modeling of a variety of energy systems that may differ in terms of spatio-temporal granularity and technological scope. Since this is the more common term, this text consistently refers to models rather than frameworks.

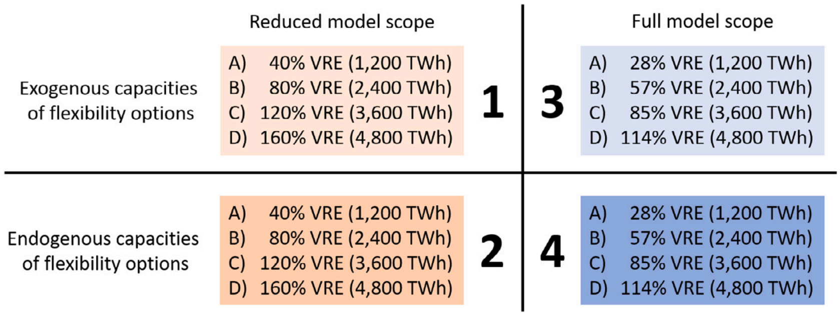

Fig. 1. Overview of the test cases considered in the model comparison. Groups 1 to 4 differ in the technology scope and consideration of endogenous capacity optimization of flexibility options. Within each group, four different sets of VRE capacities and thus amounts of renewable electricity are considered. The achievable VRE supply shares differ due to the different demands depending on the model scope.

system operation with hourly resolution. These models are applied to 16 test cases that differ substantially in their design. While one part of these cases considers a complete harmonization of the technology scope in the models, this is not the case in the other part. Thus, outcome differences can also correlate with the choice of technology scope. Regardless of the technologies considered in each test case, all models use a harmonized input data set. The focus of this study is to answer the following research questions:

1. How large are the model-related differences of optimizing power system models with complete harmonization of model scope and input data, and how are they related with the VRE share?
2. Are findings on technology deployment robust, even with different model scopes?
3. Which differences in modeling approach and technology modeling have a particularly strong impact on the composition and operation of the optimal system?

To answer these questions, model differences are first collected and categorized. On this basis, the deviations between the results are then systematically analyzed and correlated with the model differences. The results of our comparison strengthen the understanding of the effect of differences in temporally and spatially resolved power system models. This is equally helpful for developers and users of models, as well as for decision-makers in politics and industry, who use the model results.

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 their analysis. Finally, Section 4 derives the main conclusions.

2. Materials and methods

2.1. Set-up and input data of the model comparison

The basic approach in the model comparison is essentially characterized by the use of harmonized input data and stylized, yet systematic, test cases. On the data side, we build on a previous model comparison exercise, which focused on technology-specific modeling differences [7]. The harmonized input data set includes exogenous plant capacities, techno-economic parameters, and time series and is fully available at [29].

The 16 test cases differ in three characteristics: the VRE capacities, the consideration of an endogenous capacity expansion of flexibility options, and the technology scope (Fig. 1). The technology scope correlates with the degree of harmonization of the models. In the case of a reduced scope (test cases 1 and 2), the same technologies are considered in all models. In contrast, in the case of a full scope (test cases 3 and 4), there are differences between the models (Section 2.2.1). Furthermore, by additionally allowing for endogenous capacity expansion of flexibility options (test cases 2 and 4), complementary model differences to the case of exogenously given capacities (test cases 1 and 3) can become effective. Finally, by considering different sets of VRE capacities and thus supply shares, it is possible to investigate to what extent model-related differences correlate with this key parameter of power supply systems. The exact design of these three scenario dimensions is explained in more detail in the following.

The reduced systems of test cases 1 and 2 include exogenous capacities of photovoltaics (PV), wind onshore, and wind offshore as VRE technologies, and battery storage, gas turbines, and transmission lines to balance them (Fig. 2). The capacity optimization in test cases 2A–2D includes battery storage and gas turbines. Accordingly, no existing plants are assumed for these. The technology portfolio in test cases 3 and 4 includes numerous other technologies, as shown in Fig. 2. Additional controllable generation, electricity storage, demand response (DR), and sector coupling provide the system with additional flexibility to balance VRE power generation. In test cases 4A–4D, the capacities of most of these flexibility options are also optimized. Differences in the technology scope of the models in test cases 3 and 4, as well as their compensation, are presented in Section 2.2.1.

Across all test cases, a stylized system with 11 model nodes is considered. This corresponds approximately to the countries of Central Europe (Austria, Belgium, Denmark, Czech Republic, France, Germany, Italy, Luxembourg, The Netherlands, Poland, Switzerland) in terms of electricity demand, VRE potentials, and time profiles. As for sector coupling, as well as the potentials of reservoir hydro power and hydrogen cavern storage, stylized assumptions are made that do not necessarily reflect real-world conditions.

The exogenously assumed VRE capacities are designed to be theoretically sufficient to supply 40% (A), 80% (B), 120% (C), and 160% (D) of demand in the reduced system (test cases 1 and 2). However, these theoretical shares may reduce due to VRE curtailment and losses. By assuming higher demand when considering sector coupling and the full model scope (test cases 3 and 4), the theoretical VRE shares reduce to 28% (A), 57% (B), 85% (C), and 114% (D). The annual electricity demand amounts to 3020 TWh in the case of the reduced system and to 4240 TWh in the case of the full system. The VRE supply shares that can actually be realized depend on the availability and operation of the flexibility options and are analyzed in Section 3.1. The VRE capacities assumed for each model region are derived from a previous model comparison [7] and do not directly relate to actual or projected values for the associated countries.

The exogenously defined capacities of electricity storage and controllable power plants are sized to the residual peak load occurring in each country, i.e., the maximum value of the difference between demand and VRE generation. These capacities are adopted here for battery storage, hydrogen cavern storage, gas turbine power plants, and CHP plants. This leads to a structural overcapacity of flexibility in test cases 1 and 3, where at least twice the minimum capacity needed to meet the residual load is available. In order not to exacerbate this further, the consideration of reservoir hydro power plants is limited

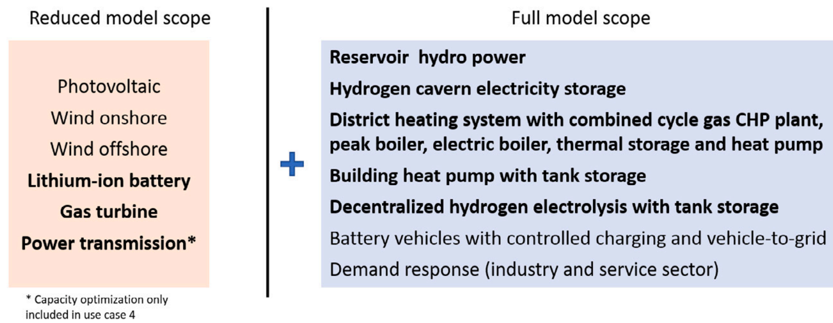

Fig. 2. Specification of the model scope in the uses cases. While test cases 1 and 2 (left side) consider only a small number of technologies, test cases 3 and 4 (right side) include numerous additional flexibility options. Bold technologies are available for capacity expansion in test cases 2 and 4 in a subset of models according to Fig. 3.

Table 1
Main characteristics of the models included in the comparison, adopted from [7]. The information on problem formulation, temporal foresight, objective and objective function refer to the model version used here and may be different in other applications.

	DIETER	E2M2	GENESYS-2	ISAaR	JMM	MarS	oemof	REMix
Modeling language	GAMS	GAMS	C++	MATLAB, PostgreSQL	GAMS	Fortran	Python	GAMS
Problem formulation	LP	LP	population-based heuristic	LP	LP	LP	LP	LP
Foresight in hours	8760	8760	1	8760	180	8760	8760	8760
Objective	min. costs	min. costs	min. costs	min. costs	min. costs	min. costs	min. costs	min. costs
Objective function	CAPEX, OPEX	CAPEX, OPEX	CAPEX, OPEX	CAPEX, OPEX	OPEX	OPEX, Lagrange multipliers	CAPEX, OPEX	CAPEX, OPEX
Documentation	[32,33]	[34,35]	[36,37]	[38–40]	[41]	[42]	[43–45]	[2,46,47]

to a selection of model regions (Austria, Switzerland, Czech Republic, France, and Italy). The capacities of the technologies for flexible sector coupling (HPs, BEVs, and hydrogen electrolyzers) were designed in [7] for a uniform electricity demand and are adopted directly. For the power grid, the capacities expected by [30] for the year 2030 are adopted. Industrial and commercial DR potentials are considered according to [31]. The assumption of a carbon dioxide (CO_2) emission price of 107 €/t is a strong driver for the use of the different flexibility options considered. All model assumptions are available in [29].

2.2. Contributing models

The model comparison includes eight well-established power system models with sector coupling. Their key characteristics are summarized in Table 1. In addition, a more detailed overview of the model properties is provided in [7], where an earlier comparison of the models with a different focus is presented.

All models consider one year of system operation in hourly resolution and minimize the system costs as specified in Section 2.3. In the configuration used here, all models are formulated as linear, non-integer problems. Since the models use a harmonized data set, differences in the results may arise for three reasons: different technology scope (Section 2.2.1), different modeling approaches (Section 2.2.2), and different technology modeling (Section 2.2.3).

2.2.1. Differences in model scope

While the models are fully harmonized in their technology scope in test cases 1 and 2, there are differences in the considered technologies in test cases 3 and 4 (Fig. 3). Capacity optimization is only relevant for cases 2 and 4. These are not considered with the JMM and MarS models. The differences in model scope result partly from whether modeling a technology is possible at all, partly from the trade-off between model complexity and solution time, and partly from model-specific choices in this model comparison.

The overview shows that there is no model pair with the same technology scope in test cases 3 and 4. Thus, not only do differences in modeling approach and technology modeling interact, but also those in model scope. To derive comparable results, substitute technologies are partially taken into account for technologies not considered (Fig. 4).

In the models that do not explicitly consider specific sector coupling technologies, such as BEVs, these are represented in a stylized way by including a respective additional, inflexible electricity demand. This ensures that the same electricity demand is met in all models. However, differences in electricity demand may arise if electricity is used for heat generation in district heating (DH) systems. In addition, there is a lower energy demand in the models without representation of CHP, since the heat demand of the corresponding DH systems is not considered elsewhere. This primarily affects the reported system costs, which do not include the provision of this heat. To consider equal capacities of controllable power plants, CHP plants are replaced by additional gas turbine power plants if they are not considered.

2.2.2. Differences in the modeling approach

With the exception of GENESYS-2 and JMM, all models optimize with perfect foresight over the overall time horizon of one year. JMM uses a rolling planning horizon for the optimization of the yearly dispatch. The year is divided into smaller periods of one week that can be solved successively to lower the complexity of the overall problem. In contrast to all other models, GENESYS-2 does not rely on a deterministic optimization, but on a population-based heuristic. Furthermore, it is designed as a dispatch model with every time step being solved independently without any foresight. In doing so, the use of technologies follows a predefined order. This order prefers local use or storage of energy before transmission. Only if there is a local surplus of VRE generation, transmission is considered. In doing so, the surplus is distributed starting with all neighboring regions and only going beyond them to more distant regions if necessary. Eventually, this leads to a more regional use of VRE. If, after the distribution of VRE, there is a shortage in demand in one region, it can request generation of other power plants from the neighbors first and then beyond.

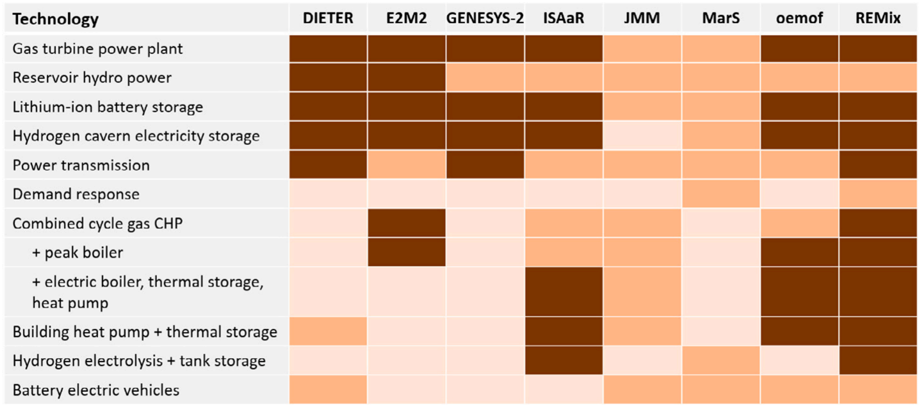

Fig. 3. Technology scope of the models applied in the comparison. The dark color indicates an endogenous capacity expansion, the middle color a consideration of exogenous capacities, and the light color a disregard of the corresponding technology. This overview does not necessarily reflect the general ability of the underlying models to consider these technologies.

Technology / Model scope	With capacity optimization	Dispatch optimization only	Not modelled
Gas turbine power plant	Green field expansion	Exogenous capacities	Not applicable
Reservoir hydro power	Green field expansion	Exogenous capacities	Not applicable
Lithium-ion battery storage	Green field expansion	Exogenous capacities	Not applicable
Hydrogen cavern electricity storage	Green field expansion	Exogenous capacities	No consideration
Power transmission	Green field expansion	Exogenous capacities	Not applicable
Demand response	Not applicable	Exogenous capacities	No consideration
Combined cycle gas CHP	Green field expansion	Exogenous capacities	Doubled gas turbine capacity (CapFix), No consideration (CapOpt)
+ peak boiler	Green field expansion	Exogenous capacities	Not applicable
+ electric boiler, thermal storage, heat pump	Green field expansion	Exogenous capacities	Only CHP and boiler considered
Building heat pump + thermal storage	Green field expansion	Exogenous capacities	Exogenous power demand profile
Hydrogen electrolysis + tank storage	Green field expansion	Exogenous capacities	Exogenous power demand profile
Battery electric vehicles	Not applicable	Exogenous capacities	Exogenous power demand profile

Fig. 4. Strategy for indirectly considering technologies that are not explicitly modeled. 'Not applicable' indicates that a certain configuration is not present in any of the models, 'No consideration' implies that no substitute for a technology is considered in the models.

2.2.3. Differences in technology modeling

The models used have numerous minor and major differences in technology modeling. The relevant ones for the following analysis include:

Power plant outages. Power plants outages have been modeled using two different approaches. In most models, outages are interpreted as a certain percentage of continuous unavailable generation capacity. This implicitly assumes an equal distribution of outages over all hours of the year. A different approach is the stochastic drawing of outages which results in partial or full unavailability of power plants, which is applied in MarS. ISAaR and oemof consider outages only for exogenous capacities, which implies that all endogenously expanded assets are available with their nominal capacity in each hour of the year.

Power plant ramping. In JMM and ISAaR, start-up processes of gas turbines and CHP plants are associated with additional fuel consumption. This leads to higher overall fuel demand and therefore to an increase in CO_2 emissions.

Representation of reservoir hydro power plants. A simplified representation of hydro power plants, where hydraulic plants are modeled as an aggregated single unit, has been used by most of the models. In GENESYS-2, natural inflows are neglected. In MarS, a cascading model is implemented which allows that natural inflows to be used multiple times. In JMM, the use of hydro reservoirs is determined by a water value which is calculated model-endogenously based on a reference filling level.

Storage and reservoir expansion. E2M2 and ISAaR use exogenously defined energy to power (E2P) ratios for some of the endogenously built storage technologies. In all other models, this ratio is optimized. E2M2 applies the exogenously defined E2P ratios considered in test case 3 to the endogenously built capacities in test case 4. For the electric energy storage units, an E2P ratio of 4 h for batteries and 400 h for hydrogen caverns is assumed. For reservoir hydro power an E2P ratio of 615 h for the pump and 400 h for the turbine is considered. In ISAaR a fixed E2P ratio of 10.4 h is assumed for the expansion of thermal energy storage (TES). Furthermore, in E2M2 the charging and discharging capacities of electricity storage must be identical, just as in ISAaR for TES. In the other models, this is only required for battery storage.

Power transmission. The most relevant difference in the representation of power transmission is the consideration of a direct current (DC) load flow approach in REMix. In contrast, all other models consider a net transfer capacity (NTC) approach, which allows higher line utilization to be realized. Transmission losses are considered in all models except MarS.

Battery electric vehicles. Differences in the modeling of BEVs particularly concern the calculation of costs. No costs are incurred for controlled charging in JMM, MarS, and oemof. In DIETER, the same specific costs are applied for each charging process independent of the timing. In REMix, costs are only incurred if there is a deviation from the exogenously specified profile of uncontrolled charging. Costs for feeding electricity back into the grid are applied in all models. The modeling in JMM also differs from the other models in that no minimum battery level is considered. In addition, it is assumed that vehicles are always fully charged before driving and are reconnected to the grid with a predefined battery level.

Thermal and hydrogen storage. There is no bypass available for building HP storage in DIETER and JMM. The same is true for hydrogen tank storage in MarS. This implies that the entire production must pass through the storage system. Consequently, the reported amounts of stored energy are larger and higher losses can occur.

CHP fuel costs. In E2M2, CHP fuel consumption is based on the equivalent electricity generation. This is calculated as the sum of electricity and heat generation, but the latter is multiplied by the power loss factor and represents the equivalent electricity generation at which the same amount of fuel is consumed for the generation of pure electricity as for the actual combined generation of electricity and heat [34].

2.3. Output indicators

The evaluation of the model comparison is essentially based on the comparison of central parameters of the technology operation. Thus, annual values of energy provision, VRE curtailment, unsupplied demand, and system costs are compared for the overall assessment area. When endogenous plant installations are considered, capacities are additionally evaluated. We use normalized indicators to allow for a better comparison of outcome deviations. In test cases 1 and 3, the reported system costs represent the variable operating costs of all assets including fuel and CO_2 costs. In test cases 2 and 4, the system costs also include the proportionate investment cost of endogenous capacity installations and their fixed operating costs, but no investment costs for exogenous capacities. In test cases 3 and 4, unsupplied demand can include heat and hydrogen in addition to electricity.

3. Results and discussion

The evaluation and analysis of the results start with the key indicator for the usage of flexibility options, which is the realized VRE share (Section 3.1). It then follows the structure of the model comparison shown in Fig. 1. Thus, the test cases with reduced, harmonized model scope without (Section 3.2) and with capacity optimization (Section 3.3) are considered first, and then those with full, heterogeneous model scope, also first without (Section 3.4) and then with capacity optimization (Section 3.5). Finally, the relation between VRE technology share and deviations in model outcomes is evaluated (Section 3.6).

3.1. Realized VRE shares

Despite the numerous model differences, we find high robustness of the VRE supply share calculated based on the used wind and PV power generation and the exogenous electricity demand (Fig. 5). In general, the deviations increase with rising VRE capacity and are higher in the cases with a full model scope than with a reduced model scope. In the test cases with the lowest VRE capacity (A), differences of less than 0.1% arise because the VRE electricity generation can be fully utilized. With reduced model scope (test cases 1 and 2), higher VRE capacities trigger differences in the realized VRE shares of up to about 5%. A much larger spread results in the case of the full model scope and the highest VRE capacities (D). Due to the different consideration of flexibility options, the difference in the achieved VRE shares reaches up to 25% there. In particular, the use of VRE for heat generation in DH contributes to this, which also enables VRE supply shares of more than 100%. Considering an endogenous capacity expansion leads to slightly lower VRE supply shares for both reduced and full model scope. However, the differences in results caused by endogenous expansion are much smaller than those between the models. Due to storage and grid losses, the VRE share in the final electricity supply may be lower and show larger differences between the models than the values shown here. This is analyzed in detail in the following.

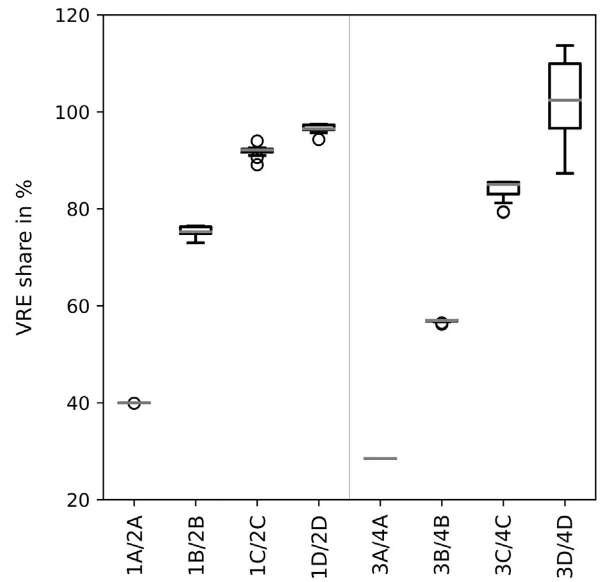

Fig. 5. Range of realized VRE supply shares in the systems with reduced model scope (test cases 1 and 2, left side) and full model scope (test cases 3 and 4, right side).

3.2. Reduced model scope without capacity expansion (test case 1)

Fig. 6 shows the key indicators for test cases 1A-1D. There, all models consider only gas turbines, battery storage, and power transmission as flexibility options.

The results reveal several deviations between the models and some clear trends. All models find a strong decrease in gas turbine usage and system costs (variable operating costs) with increasing VRE supply share, while VRE curtailment increases. The use of power storage and transmission first increases but then reaches a level of saturation at very high VRE shares when excess power generation occurs more frequently both in time and space. In addition, we find that model-related differences in system costs and power plant dispatch increase with VRE share, whereas it decreases for grid usage, storage usage, and VRE curtailment.

The most substantial deviations in model results can be clearly associated with the model differences. The largest differences are related to the use of a DC load flow approach and the assumption of no foresight over time, smaller ones due to fixed dispatch order, power plant ramping, power plant outages, and grid losses.

For the system costs, there is a very high agreement in the results across most models. Outliers arise for higher VRE shares due to deviating grid consideration (REMix) and a dispatch approach without temporal foresight (GENESYS-2). Since the system costs are essentially driven by the variable costs of gas turbine operation, a very similar pattern is observed for the latter. There, the deviation is about 25% for the outliers, and consistently less than 10% for the other models. The lowest power plant utilization and thus also the lowest costs result from neglecting the grid losses (MarS).

A structurally analogous pattern also emerges in the case of the VRE curtailment, which deviates from one another in most models only in the single-digit percentage range, with the aforementioned outliers of REMix and GENESYS-2. A more heterogeneous picture is found for the use of the electricity grid and battery storage. In both cases, very large relative deviations of over 80% arise in the case of low VRE shares (1 A), which correspond to only small absolute deviations due to the low use of the technologies. In the case of higher VRE shares, the deviation then reduces again to values below 50%, excluding the outliers even to below 20% (power grid) and 10% (battery storage). With respect to grid usage, outliers can be explained by the modeling

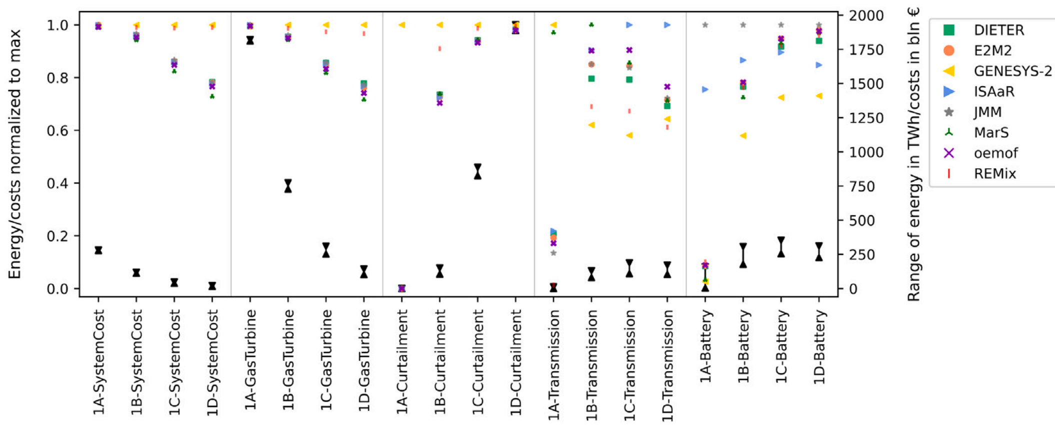

Fig. 6. Key power system operation indicators in test case 1, including system costs, power generation in gas turbines, VRE curtailment, power transmission, and battery storage output. Unsupplied energy is not reported by any of the models. The colored symbols show model-specific values for each test case and indicator normalized to the corresponding maximum according to the scale on the left y-axis. The black ranges indicate the absolute values on the scale of the right y-axis.

differences. Stochastic power plant outages are partially compensated by energy imports, which can lead to a substantial relative increase in grid usage at low VRE shares (MarS).

Using the DC load flow approach reduces the available grid capacity, which translates into lower grid utilization, which in turn must be compensated by gas turbine operation (REMix). The application of a predefined dispatch order makes the whole system more inflexible and inefficient (GENESYS-2). This results in higher system costs, power generation from gas turbines, and VRE curtailment, while at the same time reducing battery and transmission grid usage. Furthermore, more detailed modeling of gas-fired power plants involving additional higher fuel consumption for start-up processes results in a higher battery storage usage to smooth the gas turbine operation and therefore to reduce power plant start-ups (JMM, ISAaR). This effect can especially be observed at lower VRE shares where the power plant output is comparatively high.

3.3. Reduced model scope with capacity expansion (test case 2)

Test cases 2A-2D also consider a reduced and uniform technology scope, but with model endogenous optimization of battery storage and gas turbine power plant capacities. This deviation in model configuration does not change the trends in system operating parameters observed in the test cases 1A-1D for the increasing VRE supply share (Fig. 7). Thus, we see an increase in curtailment, a decrease in power plant dispatch and system costs, and an initial increase and later saturation in the use of battery storage and transmission lines. These trends emerge equally for endogenous investments in battery storage and gas turbines (Fig. 8). In contrast to test case 1, the spread of model results increases for higher VRE shares for all indicators except for VRE curtailment. This seems plausible considering the increasing endogenous plant expansion and its differences between the models.

In the endogenous addition of battery storage and gas turbines, there is a high agreement between most of the models (Fig. 8). However, individual models show substantial differences in the preferred technology and the total capacity added, which can be explained by the model differences. For example, the separate optimization of the individual time steps strongly favors the addition of gas turbines, which is why battery storage is added to a lesser extent (GENESYS-2). In contrast, when the exogenously defined E2P is lower than the optimal value, larger battery converter capacities must be provided to obtain a similar energy storage capacity as most other models (E2M2). Less pronounced is the impact of a lower usable power transmission capacity, which results in higher capacities for both storage and power plants (REMix). If a full availability of generators is considered, lower plant capacities are systematically required (oemof, ISAaR). Due to these model differences, the aggregated power generation capacity of endogenously added battery storage and gas turbine power plants differs by a maximum of 30%.

The system operation parameters show essentially the same characteristics as in the systems without endogenous capacity expansion (cf. Figs. 6 and 7). This applies equally to the trends across the models and to the deviations between the models. For example, model results are relatively similar in terms of system costs and gas turbine operation. Deviations between the models increase with the VRE share and reach a maximum of 10% in test case 2D. Outliers upwards are again the models with deviating grid modeling (REMix) and modeling methods (GENESYS-2). In contrast, a downward outlier in the costs results from the neglect of plant availability, which reduces the specific investment costs (oemof). An inflexible predefined dispatch order can cause small amounts of unsupplied energy in this test case (GENESYS-2). It is noticeable that in the case of endogenous expansion, compensation for the lower grid capacities is made to a much greater extent by batteries, which can reduce the gas turbine operation compared to the case with exogenous capacities (REMix).

3.4. Full model scope without capacity expansion (test case 3)

In the test cases 3A-3D, additional technologies are added to the power system to balance VRE power generation, as shown in Fig. 2. In doing so, all models consider only part of the full technology spectrum (Fig. 3). Nevertheless, the expected trends emerge for the main power system indicators (Fig. 9). While the costs and utilization of gas-fired power plants, here including CHP plants, decrease with the VRE share, the utilization of storage and the power grid, as well as VRE curtailment, increase. The relative deviations between the models also show different trends. While they increase with the VRE share for the costs and use of gas-fired power plants, opposite trends emerge for transmission and storage. The differences for reservoir hydro power plants are relatively constant. Remarkably, there are large differences for VRE curtailment, which can be avoided in some models across all test cases, but reaches substantial amounts in others. These substantially larger discrepancies between the results compared to the harmonized test cases can be explained by the identified model differences.

For example, neglecting electrical heat production in DH leads to much higher costs at high VRE shares, as more fuel is needed in conventional boilers (E2M2). In contrast, substantially lower system costs result if DH is not modeled, since the corresponding heat demand is not considered in the models and thus lower fuel costs are incurred (DIETER, MarS). Lower fuel costs can also be associated with the consideration of an equivalent electricity generation for calculating fuel consumption in CHP plants (E2M2).

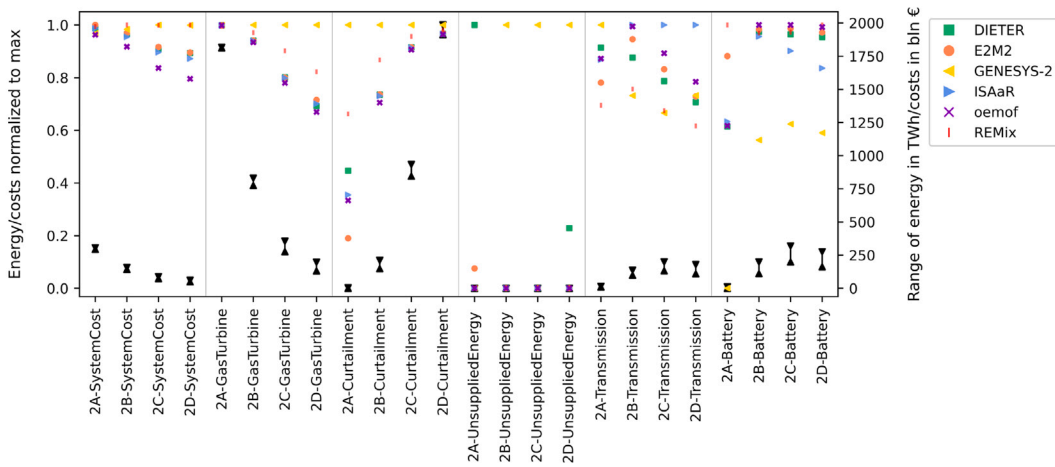

Fig. 7. Key power system operation indicators in test case 2, including system costs, power generation in gas turbines, VRE curtailment, unsupplied energy, power transmission, and battery storage output. The colored symbols show model-specific values for each test case and indicator normalized to the corresponding maximum according to the scale on the left y-axis. The black ranges indicate the absolute values on the scale of the right y-axis.

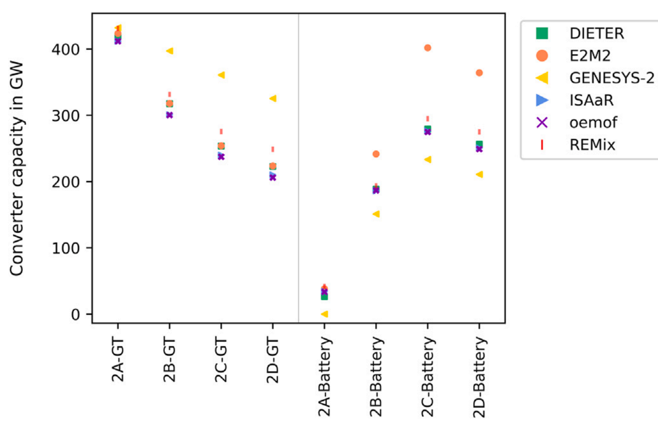

Fig. 8. Endogenous power generation capacity installations for gas turbines (GT) and battery storage converters in test case 2. In test case 1, the exogenous capacities of these technologies are 492 GW each.

The increasingly divergent power generation in gas turbine and CHP plants with higher VRE shares is closely correlated with the available other flexibility. For example, a more detailed consideration of reservoir hydro plants can increase their electricity generation at the expense of gas-fired plants with rising VRE shares until saturation where the additional available energy cannot be integrated into the system (MarS). On the other hand, reduced storage possibilities over the long term due to limited time foresight causes higher use of thermal power plants (JMM, GENESYS-2).

The use of electricity for heat generation is clearly reflected in a reduction of VRE curtailment. Long-term storage is also of great importance here, and is used heavily especially where electric heat generation is not considered (E2M2). More intensive use of long-term storage is accompanied by a decline in electricity generation from gas turbines and CHP. Where both electric heating and long-term storage are available, no curtailment is observed even with high VRE shares (ISAaR, oemof, REMix). Instead, considerably higher VRE curtailment results, if flexible sector coupling is not available and long-term storage is not operated optimally due to a lack of temporal foresight (GENESYS-2).

Although the upward trend in power grid usage with increasing VRE share is evident across all models, there is a wide dispersion in the transmitted electricity and few consistent trajectories when comparing individual models. Reason for this is a superposition of numerous model-specific effects, only a few of which can be clearly interpreted. Analogous to test cases 1 and 2, the use of a DC load flow approach reduces the amount of electricity transmitted (REMix). Furthermore, the increased range of flexible technologies diminishes the impact of stochastic modeling of outages, which resulted in increased grid usage in test case 1 (MarS).

Using a perfect foresight approach, the lower availability of flexible sector coupling options can be partially compensated by more intensive use of the power grid, reservoir hydro power, and storage (E2M2). In contrast, in models considering more sector coupling options, cavern storage is not used at VRE shares below 90% (test case 3C), and battery storage is only used to a minor extent. The exception here is the use of battery storage for the reduction of power plant startups (ISAaR, JMM).

The operation of flexible sector coupling technologies (Fig. 10) also shows clear trends. With increasing VRE share, CHP heat is replaced either by a conventional peak load boiler (E2M2), or, where possible, by electric heat generation with HP and electric boiler (ISAaR, JMM, oemof, REMix). Electrical heat supply is supplemented by TES, which provide additional flexibility. For increasing VRE shares, the use of flexible hydrogen electrolysis shows an upward trend in some models (REMix) and an almost constant trend in others (ISAaR, MarS). An increase in flexibility usage is also observed for the controlled charging of BEVs, with some models showing a saturation effect at very high VRE shares (MarS, REMix).

Even though the electrification and flexibilization of the DH supply are clearly visible in all models, there are distinct differences in implementation. This applies in particular to the use of TES, where differences of more than 70% arise. When hydrogen cavern storage is not available, TES serves as an alternative option for long-term balancing (JMM). This is accompanied by more intensive use of electric heat generation. Moreover, higher use of BEV flexibility can be observed.

There are divergent trends in the use of building TES and hydrogen tank storage, despite the rather comparable scope of the models including these technologies. An increase in storage use with the VRE share can be observed in one case for building TES (ISAaR) and in another for hydrogen tank storage (REMix). Also for BEVs, different technology modeling approaches cause substantial variations in the flexibility provided. In particular, the application of costs to a deviation from a predefined charging profile has a strong reducing effect on the use of BEV flexibility (REMix). In contrast, an increased use of this flexibility arises from neglecting a minimum battery level (JMM). An even more pronounced increase results from disregarding electrical heat generation (MarS). If flexible charging of BEVs is not possible, other flexibility options such as battery storage and hydrogen storage are used more extensively (E2M2, ISAaR).

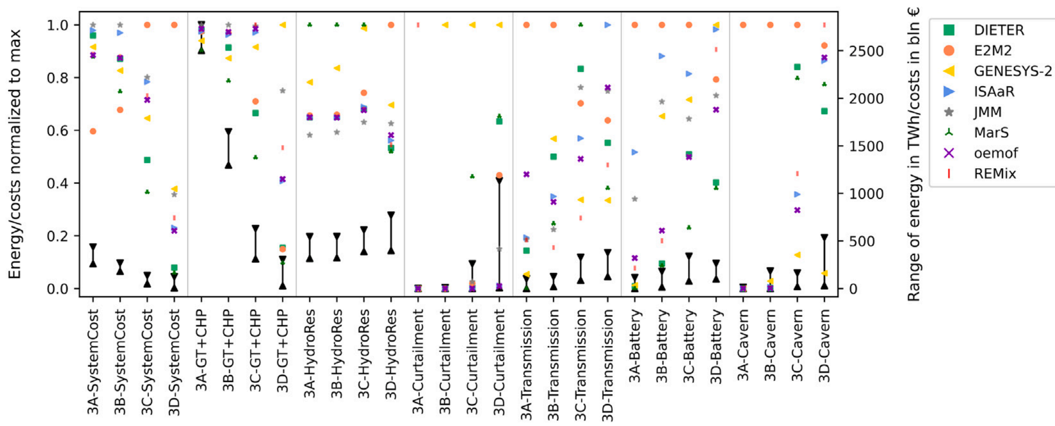

Fig. 9. Key power system operation indicators in test case 3, including system costs, power generation in gas turbines (GT) as well as CHP plants, reservoir hydro power plants, VRE curtailment, power transmission, battery storage output, and cavern storage output. Unsupplied energy is not reported by any of the models. The colored symbols show model-specific values for each test case and indicator normalized to the corresponding maximum according to the scale on the left *y*-axis. The black ranges indicate the absolute values on the scale of the right *y*-axis.

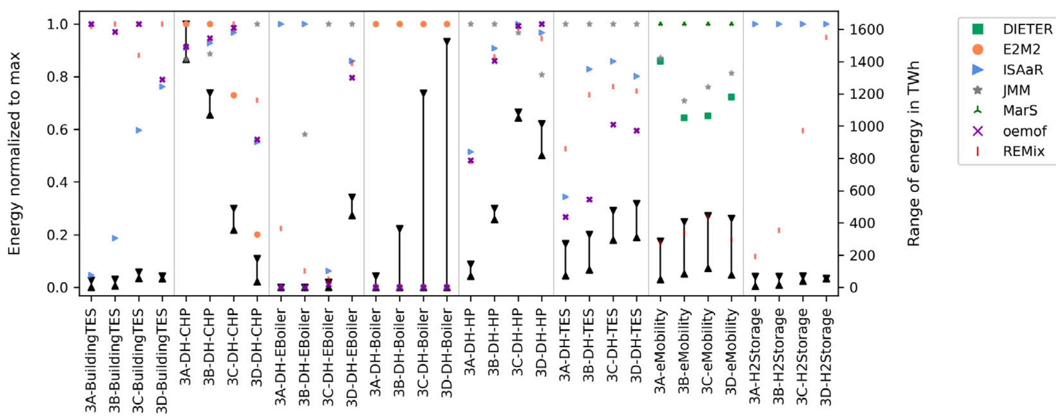

Fig. 10. Sector coupling operation indicators in test case 3, including heat production and storage, load shifting of BEVs, and hydrogen tank storage usage. The colored symbols show model-specific values for each test case and indicator normalized to the corresponding maximum according to the scale on the left *y*-axis. The black ranges indicate the absolute values on the scale of the right *y*-axis. The results of DIETER and JMM for the building TES as well as those of MarS for the hydrogen storage are not reported as these models do not consider a bypass, leading to much higher values that cannot be directly compared.

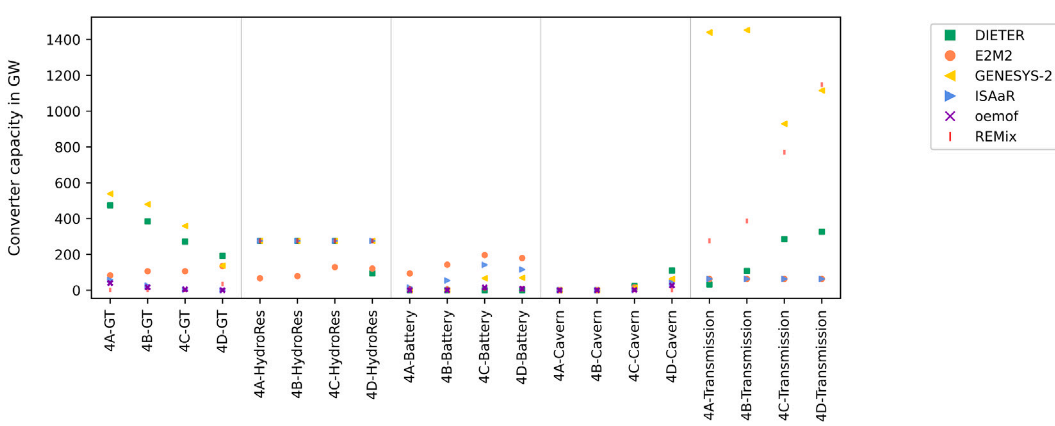

Fig. 11. Available capacities of power plants, storage, and transmission lines in test case 4. Exogenous capacities are shown here for some models in reservoir hydro power (HydroRes) and power lines (Transmission), as can be seen from the identical values in all test cases and Fig. 3.

3.5. Full model scope with capacity expansion (test case 4)

In test cases 4A-4D, the full model scope is combined with endogenous capacity expansion. This gives the models numerous additional degrees of freedom. These are used in different ways, resulting in even greater deviations of results. This can be seen, for example, in the endogenous investment decisions. It should be noted that each model has different technologies available for expansion (Fig. 3). This results in large deviations in investment decisions for power plants, electricity storage, and power lines (Fig. 11).

The substantial differences in the expansion of gas-fired power plants can be explained by the consideration of CHP. Where CHP is not

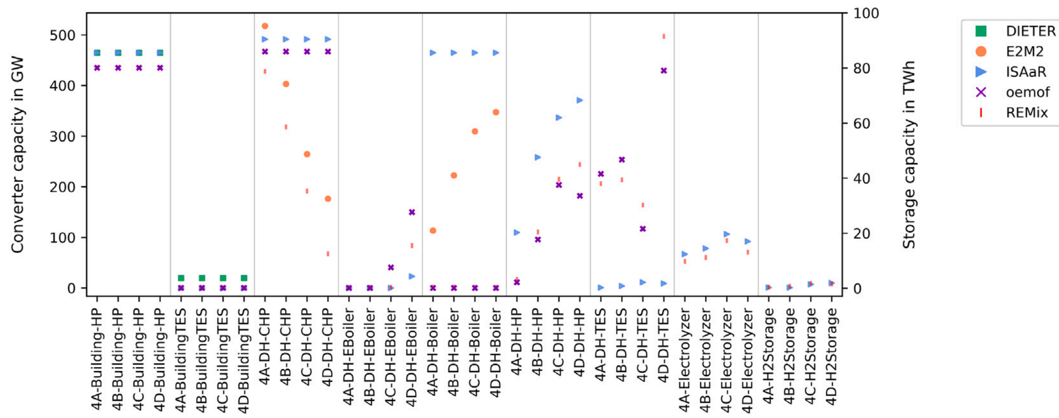

Fig. 12. Available capacities of sector coupling flexibility options in test case 4. This includes converters in the heat and hydrogen supply (left axis) and storage energy capacities for thermal and decentralized hydrogen storage (right axis). Exogenous capacities are shown here for some models according to Fig. 3.

available, much more gas turbines are needed (DIETER, GENESYS-2). In contrast, accounting for exogenous CHP capacities makes gas turbines almost completely obsolete, with small capacities only at low VRE shares (ISAaR, oemof). In the case of endogenous CHP expansion, both technologies are used (E2M2). In a combined expansion of CHP and power transmission lines, gas turbines prove to be a favorable option for residual load coverage only at very high VRE capacities (REMix).

If grid expansion is also possible, the option of expanding reservoir hydro power is used to the maximum as long as VRE generation does not exceed demand (DIETER). In contrast, if reservoir hydro power is optimized in capacity with predefined ratios and without grid expansion, lower capacities are built, which must be compensated by additional gas turbine and battery storage capacity (E2M2). Beyond that, battery storage is only added endogenously to compensate for the reduced flexibility of CHP plants if additional start-up and load change costs are considered for these (ISAaR). A small endogenous expansion of hydrogen cavern storage occurs only at higher VRE shares and when it competes with reservoir hydro power (DIETER).

An expansion of the power grid is realized where it is possible, generally increasing with the VRE share (DIETER, REMix). The use of a DC load flow approach leads to higher values for the absolute capacities (REMix). Allowing for grid expansion with a fixed dispatch order can lead to very high endogenous capacities at low VRE shares (GENESYS-2). This results from the fact that grid expansion is more attractive in the short term compared to the reduction of VRE curtailment by the addition of further storage, as there is still unmet demand in other regions. Models with deterministic optimization can balance those events using other storage when VRE shares are low but with increasing VRE share they also require higher transmission capacity. In the optimization without temporal foresight (GENESYS-2), the flexibility of storage is therefore lower.

There are also large discrepancies between the models in terms of investments in flexible sector coupling technologies (Fig. 12). Where endogenous dimensioning of CHP is possible, it is chosen to be lower with increasing VRE share (E2M2, REMix). On the heat side, this is compensated either by additional gas boilers (E2M2) or by electric heat generators (REMix). However, electrification of the DH supply is also made possible by appropriate investments if the CHP capacities are exogenously specified (ISAaR, oemof). With higher VRE shares, TES is increasingly being built for additional flexibility of DH supply (oemof, REMix). Considering a fixed E2P ratio for TES in DH networks reduces their optimal size, which in turn leads to higher HP capacities in these networks compared to the other models (ISAaR).

Decentralized TES proves to be too expensive, which is why no (ISAaR) or only a very minor expansion (oemof, REMix) takes place. With about 130 GWh, it is considerably lower than the exogenous capacities (see values for DIETER). Due to the lack of flexibility, the endogenous capacities for the associated HPs are identical in all test cases. Depending on whether the coefficient of performance of these HPs is time-variable (oemof, REMix) or constant (ISAaR), the optimal capacities are slightly different. In the latter case, higher values result, which are identical to the exogenous capacities (DIETER).

In the endogenous expansion of decentralized electrolyzers and hydrogen storage, the two models involved show similar trends. Thus, the optimal electrolyzer capacities are consistently lower than the exogenously defined ones in test case 3 (108 GW). Moreover, an increase with rising VRE share is observed in 4A–4C, as well as a slight decrease in 4D. At below 2 TWh, aggregate hydrogen tank storage capacity remains at a very low level. Thereby, an increasing trend with the VRE share can be seen in both models.

With respect to system operation, there are deviations of at least one order of magnitude between the models for most indicators and test cases (Fig. 13). Apart from a few outliers, however, the trends in the dependencies between VRE share and technology deployment of the previous test cases are confirmed. At least the larger deviations between the results can again be explained by the model differences.

The strong upward outlier in system costs results from the consideration of an endogenous capacity expansion of all conversion plants and storage facilities of sector coupling (REMix). The associated investment costs do not apply if, as a substitute, only the electricity demand of sector coupling is taken into account in other models (Fig. 4). In contrast, lower costs result where comparatively few technologies are optimized endogenously or the fuel requirement for supplying the heating networks is omitted (DIETER, GENESYS-2, ISAaR). Regarding the lower costs at low VRE shares caused by consideration of the equivalent electricity generation of the CHP plants, as well as the increased costs at high VRE shares caused by the fuel demand of the peak load boilers, the same effects result as in test case 3 (E2M2). In the usage of gas-fired power and CHP plants, there is a difference particularly in case 4D, which is caused by the endogenous grid expansion (DIETER, REMix). For reservoir hydro power plants, the most relevant effect is that a lower endogenous capacity deployment substantially reduces their electricity supply in case 4D (DIETER, E2M2). Compared to test case 3, much higher VRE curtailment is compared in some models (DIETER, E2M2, ISAaR, REMix). This generally results from incorporating the installation costs for the flexibility options into the objective function through endogenous optimization, which results in significantly lower capacities. As expected, endogenous power grid expansion leads to much higher grid usage (DIETER, GENESYS-2, REMix). The considerably lower use of electricity storage compared to test case 3 is directly related to the low endogenous installation of these facilities.

Considering an endogenous plant expansion does not result in substantial changes to the deployment of sector coupling technologies and their dependence on the VRE share (Fig. 14). Model-specific effects can

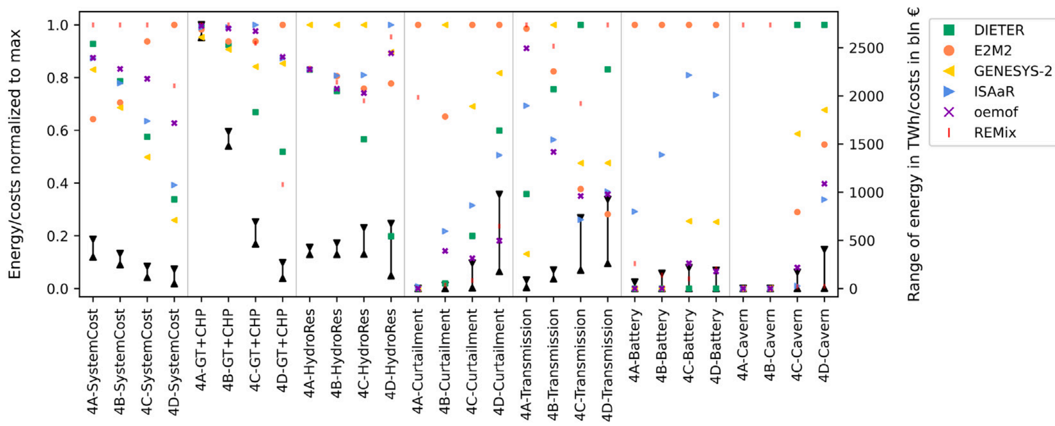

Fig. 13. Key power system operation indicators in test case 4, including system costs, power generation in gas turbines (GT) as well as CHP plants, reservoir hydro power plants, VRE curtailment, power transmission, battery storage output, and cavern storage output. Very small amounts of unsupplied energy only occur for the GENESYS-2 model in test case 4C. The colored symbols show model-specific values for each test case and indicator normalized to the corresponding maximum according to the scale on the left *y*-axis. The black ranges indicate the absolute values on the scale of the right *y*-axis.

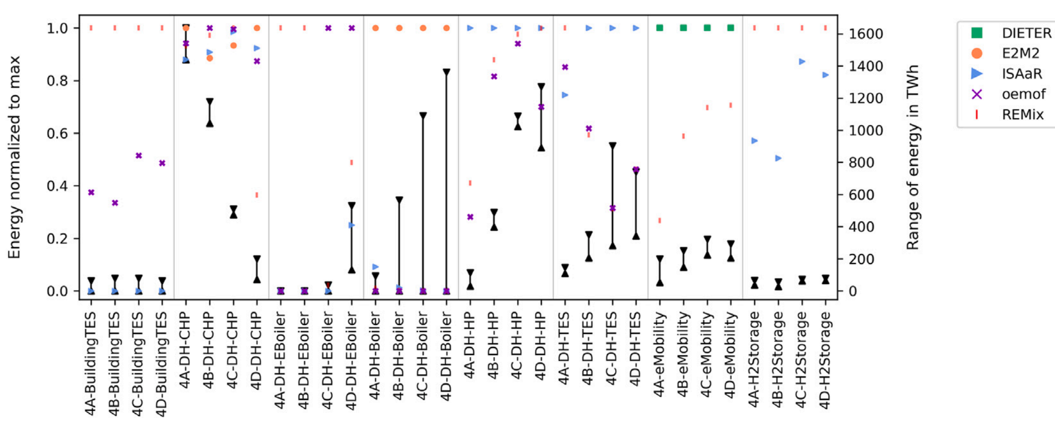

Fig. 14. Sector coupling operation indicators in test case 4, including heat production and storage, load shifting of BEVs, and hydrogen tank storage usage. The colored symbols show model-specific values for each test case and indicator normalized to the corresponding maximum according to the scale on the left *y*-axis. The black ranges indicate the absolute values on the scale of the right *y*-axis. The results of DIETER for the building TES are not reported as these include all heat supplied due to the missing storage bypass.

also be attributed to the same causes. Substantial differences to the cases with exogenous capacities (Section 3.4) are only observed for TES and hydrogen tank storage. This results from the different endogenous dimensioning in the models (Fig. 12).

3.6. Regional effects

The consideration of a multi-node system allows the analysis of the dependence between model deviations and VRE supply share (Fig. 15). This is examined using the VRE curtailment as an example. The average deviation from the median over all models is considered. Since there are numerous scenarios and models in test cases 3 and 4, in which no VRE curtailment occurs, the analysis only includes test cases 1 and 2.

The results show mean deviations from the median of a maximum of 20%. There is a clear correlation with the distribution between wind and solar energy. While the median deviation in Poland, where 77% of the VRE electricity generation comes from wind energy and only 23% from PV, is 0%, the highest values of more than 15% are found in Belgium, Luxembourg, and Switzerland, where PV accounts for at least 60% of the generation in each case. This suggests that the present model differences have a larger effect at high PV shares.

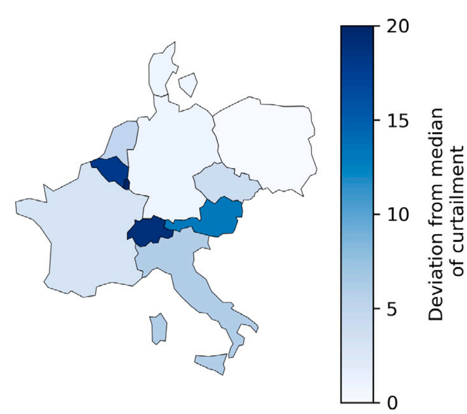

Fig. 15. Regional differences in model deviations. Shown are the averaged deviations from the median of the VRE curtailment, averaged over all test cases 1 and 2. As explained in Section 2.1, the power systems modeled here are highly stylized and correspond only in individual parameters to those of the countries depicted.

4. Conclusions

In a structured comparison, we quantify model-related deviations in the outcomes of eight power system models with sector coupling for 16 test cases. Our model results are very similar in the case of a completely harmonized and reduced model scope with pure dispatch optimization of VRE, electricity storage, power transmission, and gas turbine power plants (test case 1). Here, the considered indicators of technology operation mostly deviate by less than 40%. For most of the models, the differences are even less than 20%. These general findings do not change if endogenous capacity optimization of gas turbines and storage is also considered (test case 2). In general, the relatively high agreement of the results suggests high robustness of the models, and can be taken as a sign of their validity.

As expected, the consideration of more comprehensive and heterogenous technology scopes leads to considerably larger deviations in results than in the case with fully harmonized models. Even with a completely exogenous specification of plant capacities (test case 3), there are deviations in technology operation of more than one order of magnitude in some cases, and even higher for some outliers. However, there are also subgroups of models with similar results, especially when the model scope differs only slightly. The ranges of results are even higher when additional model degrees of freedom are available by considering endogenous plant expansions of further flexibility options (test case 4). Nevertheless, it is possible to compare the results in a meaningful way and to explain the observed deviations by the identified model differences. While the deviations between the models in the use of the individual flexibility options converge with increasing VRE share for the reduced model scope and exogenous capacities, the opposite picture emerges for the comprehensive model scope and capacity optimization.

Our analysis shows that different model scopes and modeling approaches have substantially larger impacts than the identified differences in technology modeling. With regard to the modeling approach, this relates here in particular to the use of a heuristic approach with a predefined technology dispatch order, and to a lower extent to the use of a rolling time horizon. In terms of technology scope, neglecting the flexible supply of district heating with CHP and heat pumps as well as flexible charging of battery vehicles have a substantial impact on results. As regards technology modeling, differences for reservoir hydro power plants and the power grid lead to the largest deviations in plant deployment, whereas considering fixed storage designs and neglecting power plant outages lead to the largest deviations in investments.

Despite the sometimes very large deviations between the model results, robust effects emerge with regard to the use of technologies and their dependence on the VRE share. Consistent outcomes include, for example, the flexibility of sector coupling, which is used across all models to the extent that it is considered. This particularly concerns the flexibility of vehicle charging and the partially electric heat production in district heating. Moreover, we show that even with theoretical VRE shares well above 100% of demand and a broad portfolio of flexibility options, controllable power plants or CHP plants are not abandoned completely. The increased use of the power grid and long-term storage at higher shares of VRE is also consistent in all models. In contrast, there is no clear picture for the use of stationary batteries, as these are replaced by flexible sector coupling in some models. Nevertheless, whether and how individual flexibility options are used in the models is closely linked to the questions of which other technologies are considered and how they are modeled. This must be taken into account in the interpretation and evaluation of model results.

The appearance of robust results in technology usage suggests that the models can be used to address the same issues despite their differences in detail. Still, the specific characteristics and specializations should be taken into account when selecting the model, as they can have a significant impact on the results. The strengths of the model used in each case should match the application. For example, for the analysis of integrated future energy systems, it is helpful to have a comprehensive representation of sector coupling, whereas, for the evaluation of security of supply, a focus on controllable and stabilizing plants is more suitable. As an alternative to an application-oriented use of different models, integration of the different modeling approaches into one model would be feasible. Joint use of particularly detailed technology modeling approaches would probably lead to unacceptable solution times of the mathematical problem in a spatially and temporally resolved model. With regard to the models compared here, this for example includes the detailed representation of thermal power plants, reservoir hydro cascades, and demand response. Therefore, flexible scalability of the technological complexity would be desirable for an integrated model. Thus, an adaptation to the respective application could be done. The consideration of different approaches to technology modeling in one model offers the potential for complementary comparisons that focus on different versions of one model.

With regard to the method used, our analysis shows that the use of harmonized input data and profound analysis of model properties allow the association of key outcome deviations with model differences. Thus, the effects of different modeling approaches can be captured and quantified. In addition, the effects of considering or not considering individual flexibility options on the operation of the modeled system can be analyzed. Most of the analyzed effects can be observed in several models. This suggests that these findings on the effect of model differences can also be applied to other models based on a cost-minimization approach. This gives other modelers and users of model-based energy scenarios the possibility to better interpret the results. To further strengthen the understanding of model robustness, future comparison studies should consider larger sets with identical model scope even for a full analysis of all flexibility options. While models that only cover the traditional power sector are easier to harmonize, comparative studies of models with numerous flexibility and sector coupling technologies are more challenging. This concerns the harmonization of both the technology scope and of cross-sectoral input data. Here we see a promising field for future research on more detailed model comparisons.

By considering different VRE shares and several regions in the scenarios, a broad spectrum of supply structures is taken into account. This increases the possibility of transferring general findings to more realistic scenarios. Nevertheless, follow-up studies should additionally quantify the impact of model choices on outcomes of realistic transformation scenarios.

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:** Investigation, Writing – original draft, Writing – review & editing, Funding acquisition. **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. **Felix Gaumnitz:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Jonas van Ouwerkerk:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Jennifer Mikurda:** Methodology, Software, Validation, Formal analysis, Investigation, Writing- Original draft preparation, Writing – review & editing. **Laura Torralba-Díaz:** Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

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 model output data, the input data, and the data template used are available on https://doi.org/10.5281/zenodo.5802178.

Acknowledgments

The research for this paper was performed within the FlexMex project supported by the German Federal Ministry of Economic Affairs and Climate Action under grant number 03ET4077A-G. 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. Furthermore, we thank Alexander Fehler, Alexander Zerrahn, Christine Krüger and Tomke Janßen for their contribution to model preparation and discussion on the models.

References

[1] Blanco H, Faaij A. A review at the role of storage in energy systems with a focus on Power to Gas and long-term storage. Renew Sustain Energy Rev 2018;81:1049–86. http://dx.doi.org/10.1016/j.rser.2017.07.062.

[2] Gils HC, Gardian H, Schmugge J. Interaction of hydrogen infrastructures with other sector coupling options towards a zero-emission energy system in Germany. Renew Energy 2021;180:140–56. http://dx.doi.org/10.1016/j.renene.2021.08.016.

[3] Richardson DB. Electric vehicles and the electric grid: A review of modeling approaches, impacts, and renewable energy integration. Renew Sustain Energy Rev 2013;19:247–54. http://dx.doi.org/10.1016/j.rser.2012.11.042.

[4] Bloess A, Schill W-P, Zerrahn A. Power-to-heat for renewable energy integration: A review of technologies, modeling approaches, and flexibility potentials. Appl Energy 2018;212:1611–26. http://dx.doi.org/10.1016/j.apenergy.2017.12.073.

[5] Hanley ES, Deane JP, Ó Gallachóir BP. The role of hydrogen in low carbon energy futures – A review of existing perspectives. Renew Sustain Energy Rev 2018;82:3027–45. http://dx.doi.org/10.1016/j.rser.2017.10.034.

[6] Kwakkel JH, Walker WE, Marchau VA. Classifying and communicating uncertainties in model-based policy analysis. Int J Technol Policy Manag 2010;10(4):299–315. http://dx.doi.org/10.1504/IJTPM.2010.036918.

[7] Gils HC, Gardian H, Kittel M, Schill W-P, Zerrahn A, Murmann A, et al. Modeling flexibility in energy systems - comparison of power sector models based on simplified test cases. Renew Sustain Energy Rev 2022;158:111995. http://dx.doi.org/10.1016/j.rser.2021.111995.

[8] Gacitua L, Gallegos P, Henriquez-Auba R, Lorca A, Negrete-Pincetic M, Olivares D, 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.

[9] Lopion P, Markewitz P, Robinius M, Stolten D. 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.

[10] Savvidis G, Siala K, Weissbart C, Schmidt L, Borggrefe F, Kumar S, 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.

[11] Prina MG, Manzolini G, Moser D, Nastasi B, Sparber W. 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.

[12] Fattahi A, Sijm J, Faaij A. 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.

[13] Klemm C, Vennemann P. 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.

[14] Sugiyama M, Fujimori S, Wada K, Endo S, Fujii Y, Komiyama R, et al. Japan's long-term climate mitigation policy: Multi-model assessment and sectoral challenges. Energy 2019;167:1120–31. http://dx.doi.org/10.1016/j.energy.2018.10.091.

[15] Bistline JE, Brown M, Siddiqui SA, Vaillancourt K. Electric sector impacts of renewable policy coordination: A multi-model study of the North American energy system. Energy Policy 2020;145:111707. http://dx.doi.org/10.1016/j.enpol.2020.111707.

[16] Huntington HG, Bhargava A, Daniels D, Weyant JP, Avraam C, Bistline J, et al. Key findings from the core North American scenarios in the EMF34 intermodel comparison. Energy Policy 2020;144:111599. http://dx.doi.org/10.1016/j.enpol.2020.111599.

[17] Giarola S, Molar-Cruz A, Vaillancourt K, Bahn O, Sarmiento L, Hawkes A, et al. The role of energy storage in the uptake of renewable energy: A model comparison approach. Energy Policy 2021;151:112159. http://dx.doi.org/10.1016/j.enpol.2021.112159.

[18] Gils HC, Pregger T, Flachsbarth F, Jentsch M, Dierstein C. 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.

[19] Siala K, Mier M, Schmidt L, Torralba-Díaz L, Sheykkha S, Savvidis G. Which model features matter? An experimental approach to evaluate power market modeling choices. 245, 2022, p. 123301. http://dx.doi.org/10.1016/j.energy.2022.123301,

[20] Mai T, Bistline J, Sun Y, Cole W, Marcy C, Namovicz C, et al. The role of input assumptions and model structures in projections of variable renewable energy: A multi-model perspective of the U.S. electricity system. Energy Econ 2018;76:313–24. http://dx.doi.org/10.1016/j.eneco.2018.10.019.

[21] Ruhnau O, Bucksteeg M, Ritter D, Schmitz R, Böttger D, Koch M, et al. Why electricity market models yield different results: Carbon pricing in a model-comparison experiment. Renew Sustain Energy Rev 2022;153:111701. http://dx.doi.org/10.1016/j.rser.2021.111701.

[22] Misconel S, Leisen R, Mikurda J, Zimmermann F, Fraunholz C, Fichtner W, et al. Systematic comparison of high-resolution electricity system modeling approaches focusing on investment, dispatch and generation adequacy. Renew Sustain Energy Rev 2022;153:111785. http://dx.doi.org/10.1016/j.rser.2021.111785.

[23] Pietzcker RC, Ueckerdt F, Carrara S, de Boer HS, Després J, Fujimori S, et al. System integration of wind and solar power in integrated assessment models: A cross-model evaluation of new approaches. Energy Econ 2017;64:583–99. http://dx.doi.org/10.1016/j.eneco.2016.11.018.

[24] Bosetti V, Marangoni G, Borgonovo E, Diaz Anadon L, Barron R, McJeon HC, et al. Sensitivity to energy technology costs: A multi-model comparison analysis. Energy Policy 2015;80:244–63. http://dx.doi.org/10.1016/j.enpol.2014.12.012.

[25] Wilkerson JT, Leibowicz BD, Turner DD, Weyant JP. Comparison of integrated assessment models: Carbon price impacts on U.S. energy. Energy Policy 2015;76:18–31. http://dx.doi.org/10.1016/j.enpol.2014.10.011.

[26] Aldy J, Pizer W, Tavoni M, Reis LA, Akimoto K, Blanford G, et al. Economic tools to promote transparency and comparability in the Paris agreement. Nature Clim Change 2016;6:1000–4. http://dx.doi.org/10.1038/nclimate3106.

[27] Krey V, Guo F, Kolp P, Zhou W, Schaeffer R, Awasthy A, et al. Looking under the hood: A comparison of techno-economic assumptions across national and global integrated assessment models. Energy 2019;172:1254–67. http://dx.doi.org/10.1016/j.energy.2018.12.131.

[28] Giarola S, Mittal S, Vielle M, Perdana S, Campagnolo L, Delpiazzo E, et al. Challenges in the harmonisation of global integrated assessment models: A comprehensive methodology to reduce model response heterogeneity. Sci Total Environ 2021;783:146861. http://dx.doi.org/10.1016/j.scitotenv.2021.146861.

[29] Hedda G, Gils HC, Kittel M, Murmann A, Launer J, Gaumnitz F, et al. Model input and output data of the FlexMex model comparison. Zenodo; 2021, http://dx.doi.org/10.5281/zenodo.5802178.

[30] ENTSO-E: TYNDP - Market modelling data. 2016, [cited 22.01.2022]. URL https://www.entsoe.eu/Documents/TYNDP%20documents/TYNDP%202016/rgips/TYNDP2016%20market%20modelling%20data.xlsx.

[31] Gils HC. 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.

[32] Zerrahn A, Schill W-P. 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.

[33] Gaete-Morales C, Kittel M, Roth A, Schill W-P. 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.

[34] Sun N. 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.

[35] Torralba-Díaz L, Schimeczek C, Reeg M, Savvidis G, Deissenroth-Uhrig M, Guthoff F, 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.

[36] Bußar C. 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.

[37] Siemonsmeier M, Bracht N, Bußar C. 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.

[38] Böing F, Murmann A, Pellinger C, Kigle S. ISAaR - Integrated simulation model for unit dispatch and expansion with regionalization. Tech. rep., FfE; 2019, URL https://www.ffe.de/en/isaar.

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

[40] Böing F. Cross-sector assessment of CO_2 abatement measures and their impact on the transmission grid [Ph.D. thesis], Technical University of Munich; 2020, https://mediatum.ub.tum.de/1539536.

[41] Meibom P, Barth R, Hasche B, Brand H, Weber C, O'Malley M. 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.

[42] Drees T. 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.

[43] Hilpert S, Kaldemeyer C, Krien U, Günther S, Wingenbach C, Pleßmann G. 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.

[44] Krien U, Schönfeldt P, Launer J, Hilpert S, Kaldemeyer C, Pleßmann G. 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.

[45] oemof.solph documentation. URL https://oemof-solph.readthedocs.io/en/latest/.

[46] Gils HC, Scholz Y, Pregger T, de Tena DL, Heide D. 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.

[47] Gils HC, Simon S. 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.