

Decarbonization Pathways of Worldwide Energy Systems – Definition and Modeling of Archetypes

Martin Kueppers ¹, Stephany Nicole Paredes Pineda ¹, Michael Metzger ¹, Matthias Huber ¹, Simon Paulus ¹, Hans Joerg Heger ¹, and Stefan Niessen ¹

¹Affiliation not available

October 30, 2023

Abstract

This paper includes an approach to summarize energy systems of countries worldwide in energy system archetypes. These archetypes are generated by a clustering algorithm and modeled in an energy system optimization model by standardized rules. The results show that archetypes provide an opportunity to compare decarbonization pathways of different countries on a global level. Furthermore, they can be used to simplify data collection and elaborate the potential of new technologies in different energy systems.

Decarbonization Pathways of Worldwide Energy Systems – Definition and Modeling of Archetypes

Martin Kueppers^{a,b,*}, Stephany Nicole Paredes Pineda^a, Michael Metzger^a, Matthias Huber^a, Simon Paulus^a, Hans-Joerg Heger^a, Stefan Niessen^{a, b}

^a Siemens AG, Corporate Technology, Research in Energy and Electronics, Otto-Hahn Ring 6, Munich 81739, Germany

^b Technology and Economics of Multimodal Energy Systems, Technical University of Darmstadt, Landgraf-Georg-Straße 4, 64283 Darmstadt, Germany

Abstract:

Energy system models help to find the optimal technology mixes for decarbonization strategies in countries worldwide. To reduce the modeling effort and analyze as many countries as possible, this paper proposes a novel approach of energy system archetypes which can be directly evaluated. These archetypes classify similar countries worldwide independently from their geographic location. Advantages of this idea are the setup of a transferable global database allowing for data reconstruction between countries, market size estimations, and the ability to compare peer countries facing similar challenges. To enable such modeling, a framework is developed in which the archetypes are defined, standardized modeling rules are developed, and the results are evaluated for validation. As a result, the presented clustering method improves a country classification by 44% compared to a pure geographic classification. The modeling results state the importance of balancing technologies for the daily cycle of photovoltaic generation and the importance of flexibility in future decarbonized energy systems. Overall, the results confirm that archetypes are an adequate approach to derive the set of solutions for the decarbonization of worldwide countries.

Keywords:

energy system modeling, archetype, decarbonization, country energy systems, clustering

1. Introduction

1.1 Motivation

Climate change is a major challenge which is currently faced by countries around the world. In the Paris Agreement, which was adopted during the 21st Conference of the Parties (COP21), all nations agreed on the common goal to keep the temperature rise below 2 degrees and preferably below 1.5 degrees Celsius above pre-industrial level [1].

Four years later, a report published by the “United Nations Environment Programme” indicates that global emissions have been rising by around 1.5% per year during the last decade. This trend can be attributed mostly to CO₂ emissions from energy sector and industry which rose by 2% in 2018 alone. As a result, the report emphasizes the importance of decarbonizing the energy sector and electrifying end uses for a successful transition to a low carbon society [2]. Current policies and national determined contributions (NDC) are clearly not enough to meet these ranges. Consequently, these policies must be continuously analyzed worldwide to reach the agreed goals of the Paris Agreement.

Since there is no global “one-fits-all” solution to determine the decarbonization pathways of all countries to reach the climate goals, a more detailed analysis of this transition and the required technologies is necessary. Modeling the energy system can support this analysis so that the current state of the energy system as well as general conditions within a country, such as renewable supply conditions or the availability of resources, can be considered. Furthermore, the future energy mix and selection of technologies for certain scenarios, as e.g. a decarbonization scenario, can be analyzed by such models.

Taking models as means to tackle the global decarbonization challenge, its application on as many UN member states as possible leads to two questions: How can as many countries as possible be modeled in

* Corresponding author, E-Mail address: martin.kueppers@siemens.com (M. Kueppers)

a comparable and transferable way if their current energy systems are different? Can similarities between countries help to anticipate future developments, e.g. regarding their status on a decarbonization pathway or with respect to favored future technologies? This paper aims at answering those two major questions by finding a standardized modeling approach and apply it to an energy system model.

1.2 Analysis of energy system models with global application

In general, there are various energy system models available which focus on different applications, scales of energy systems or degrees of details. Some review articles cover the broad range of available energy system models [3-6], while others focus on specific model classes as agent-based models [7] or statistical models [8].

Concerning the topic of analyzing national energy systems regarding their medium- and long-term development strategies, the model class of energy system optimization models is often used in the mentioned reviews [3-6]. It is also suitable to support political decision processes [3] for decarbonization. Therefore, this model class is selected for our application. A few studies focus on this model class and its characteristics in more detail: [3] elaborates, why this class is suitable for national decarbonization strategies, [4] describes the combination of an optimization model with a scenario and an operational power simulation model, and [5] compares the advantages of an optimization model with a detailed power system model. While the three mentioned articles either compare many different models in the defined class or the differences with similar model classes, [6] focuses on the general description of such a model. Therein the advantages of doing holistic energy system analyses including multiple sectors are explained, different criteria which must be considered for modeling are named and a typical mathematical problem formulation is provided.



Figure 1. General structure of the energy system modeling process adapted from [9]

All optimization models focusing on a national level have a similar modeling process, which is described in Figure 1. The four major steps are collecting the input data, the model itself, generating raw output data, and the interpretation of results [9]. Concerning the model itself as the central part of this process, frequently mentioned models in the above described class are MESSAGE, TIMES or OSeMOSYS [3-5,10,11]. Nevertheless, the model is not the only part of the modeling process. The input data must be gathered and processed to be applicable in the model, which is considered a big challenge [6, 12].

Since the goal is to analyze the decarbonization transitions of as many UN countries as possible, both, the model and the input data, are required to be globally applicable and available on a comparable basis. Comparing the three above mentioned models, all have been applied in global models [13-15]. In [16], global pathways are analyzed to evaluate the feasibility of a global power grid, and in [17,18] an energy system model is applied to model a decarbonization transition in a detailed level. However, all the above-mentioned studies aggregate or disaggregate the data based on their geographic location. They assume that countries in the same geographic region have similar energy systems which is not necessarily the case. In order to separate the idea of analyzing countries in a global context from the continent, sub-continent or region they are located, the idea of energy system archetypes was introduced [19]. Thereby, the characteristics of each country and similarities between countries can be analyzed, such as the difference between a country like Norway, which is dominated by hydropower generation, and Saudi-Arabia, whose energy system is dominated by oil and gas [20,21]. In [22], this archetype concept is described as follows: “Archetype analysis investigates recurrent patterns of the phenomenon of interest at an intermediate level of abstraction to identify multiple models that explain the phenomenon under particular conditions”. This idea is further pursued in the present paper by extending the archetypes generated in [19] and applying them in an energy system model. The archetypes in general and the results of their application in an energy system model have several advantages:

- For modelers: data can be transferred between similar countries in case of lack of data. The data just needs to be scaled by predefined factors. This allows for a transferrable database.
- For policy makers: the similarity of different countries can be analyzed in the current system as well as when comparing future decarbonization pathways. Consequently, policies can be designed based on this comparison, e.g. promoting further renewable energy sources.
- For technology providers: the potential of certain technologies can be estimated with a simple first analysis allowing for an initial market potential study. The most interesting countries can later be assessed in a more detailed manner.

The idea of summarizing countries by considering different characteristics has already been applied to energy-specific topics. By comparing nine categories which describe the energy productivity, [23] analyzes 39 countries and builds three clusters with a K-means clustering algorithm. Another clustering approach, a model-based clustering, is applied in [24] to the countries in the European Union evaluating their historical development of the energy mix. [25] uses 59 indicators to cluster 11,131 municipalities within Germany as a preparation to model decentral energy systems using the Ward's method. In [26] countries are not directly clustered, but typical countries that describe different types of energy systems are analyzed. The countries are selected according to their power system transformation process by considering different phases, general conditions, and policy measures. In general, the idea of global energy system archetypes has already been published before in [27]. These archetypes were generated by a qualitative hierarchical analysis of 40 countries in 5 different classes. Summarizing all mentioned sources, there is no quantitative approach, which describes the energy system of a country and which can be applied to all UN member countries globally in a bottom-up approach, grouping countries to archetypes. Furthermore, the link to an energy system model and the standardized evaluation of its results are also missing in the analyses conducted so far. The generated archetypes can be “understood as mental representations of relationships between attributes and processes that characterize systems” [28].

1.3 Goals

Summarizing the presented idea, the goal of our approach in this paper is to determine challenges and technological solutions in the global decarbonization process by modeling energy system archetypes. The archetypes must be defined in a first step by a classification methodology which summarizes countries worldwide independent from their geographic location. For validation purposes, the modeling results of the archetype can be compared to country results to analyze, how well these countries are represented by its archetype in the energy system model.

In section 2, the overall framework for the archetype assessment is presented by describing each step of the process. Subsequently in section 3, the results are presented for each step. Finally, the conclusions and learnings during the modeling process are outlined in section 4. Therein, the suitability of energy system archetypes to summarize country energy systems and analyze their decarbonization strategies is assessed.

2. Methodology

In the following section, the framework which is used to generate, model and analyze the energy system archetypes is described. Based on the modeling phases described in Figure 1, Figure 2 provides an overview of the overall framework. The three methodological steps are described in section 2.1, 2.2, and 2.3 accordingly.

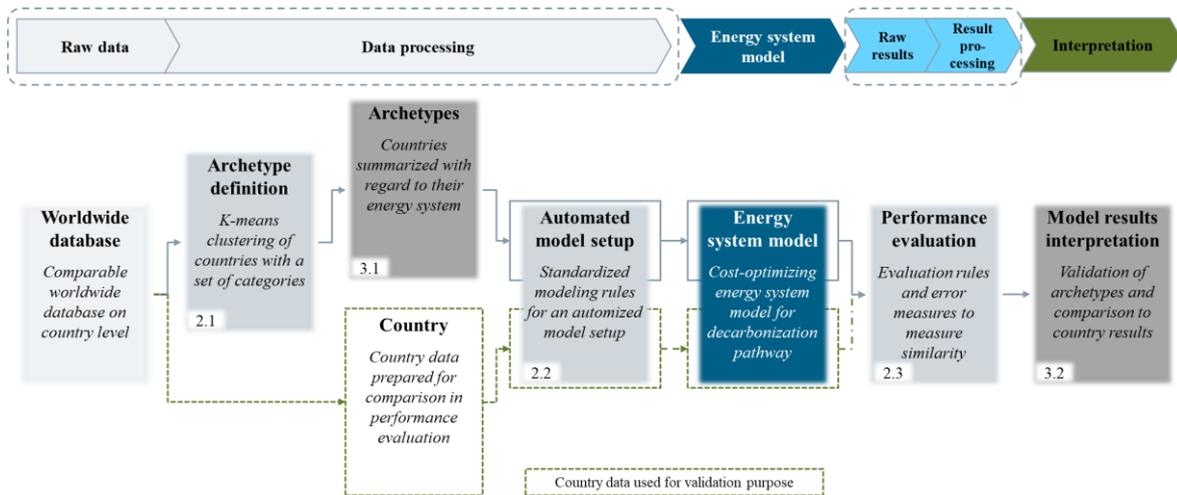


Figure 2. Overview developed framework for archetype generation and application to energy system model

The process starts with the collection of raw data. In our application, the raw data is required to be a worldwide database on country level for as many categories as possible. The energy system of a country can not only be characterized by energy-related data as it is also influenced by socio-economic and geographical characteristics [26]. Consequently, the database consists of these three categories. In these categories, data is collected for all 193 UN member countries. A more detailed description of the data acquisition and criteria for the selection of sources can be found in [19].

In the data processing step, the database is used as a basis for the clustering algorithm (see section 2.1) to generate the energy system archetypes. In parallel, it can also be used to derive country information directly. Both, the archetype and the country data, can be processed to set up an energy system model. Therefore, standardized modeling assumptions and rules are needed (see section 2.2). A linear optimization algorithm minimizing the overall system costs is applied in these models to achieve a cost-effective decarbonization pathway [6].

The Energy System Development Plan (ESDP) solves this problem and simultaneously optimizes the operation schedule and capacities of the energy system [29]. It uses a multi-modal energy approach, meaning that coupling between different energy sectors can be modeled. The model is applied to find the optimal system for a target year while also considering operation costs and determining the ideal dispatch schedule. To completely define the optimization problem, the evaluated technologies must be specified by technical and financial assumptions. A more detailed description of the model can be found in [30].

In the modeling process, the archetype model is the major focus of this paper. Additionally, the country model can be used for validation purposes: the modeling rules can be validated by a test of historical data and the archetypes can be validated by comparing the modeling results of a selected archetype with the results of countries within this archetype. For both, the country and the archetype model, the same modeling rules are applied. This step is performed in section 2.3 by gathering the raw results in a standardized way and processing the results. Finally, in the interpretation phase, patterns and conclusions are derived based on the modeling results.

2.1 Archetype definition

For the methodological characterization of archetypes, there are three required steps defined and described in [28]: (1) a set of attributes, (2) a mathematical model, and (3) a definition of the domain, in which the archetypes are valid. In the following sections these steps are described applied to the energy system context. A more detailed description of the applied process, including a formulation of the mathematical problem, can be found in [19].

2.1.1 Set of attributes – Selection of data categories from database

Although the database consists of many data categories, only the most relevant ones are selected as variables to describe the energy system of a country. In short, the four most important selection criteria are: (1) variables must not be duplicate or similar, (2) variables need to be comparable, (3) variables must be available for at least 85% of the UN countries and (4) variables must be comparative or normalized [19]. Additionally, data which is directly needed for the setup of an energy system model must also be included in the clustering. Overall, the selected variables should respect all elements of an energy system by also taking socio-economic and climatic / geographic data into account [26].

2.1.2 Mathematical model – clustering algorithm

Cluster analysis is mentioned as a suitable methodology to classify data into archetypes [28]. Clustering methods are in general applied to compress data from a large dataset into simplified statements [31]. Therefore, we chose a clustering method to generate the energy system archetypes.

To determine an appropriate clustering method, the following criteria must be met: the methodology must handle large datasets, directly assign each country to one cluster and consider all data categories bottom-up and simultaneously. The K-means algorithm is chosen, as it complies with the defined criteria [31]. This algorithm has also already proven to be suitable for the classification of countries by clustering different categories [23] and is also mentioned as a suitable algorithm to generate archetypes [28].

In order to generate reasonable and stable results, the standard K-means algorithm is extended through the addition several features, whose suitability have already been proven in general reviews and other applications [31,32]:

- The data categories are all normalized into a range of [0,1] before starting the clustering.
- The initial partition of K-means, which is decisive for the result, is improved by the K-means++ approach [33].
- The algorithm is repeated iteratively so that the final clustering result, which is generated heuristically, is chosen from the best solution found in all iterations.

Another decisive criterion when applying K-means is the number of clusters K . Both, literature about cluster analysis and about archetype generation, state that there is no general approach to determine the number of clusters [28,31]. While the improvement of the Euclidean distance within the clusters is a good criterion to consider, the number of clusters also depends on the application of the clusters. Next to the data-driven distance analysis, we also considered other studies which summarize countries globally for energy system models [19]. Lastly, the number of clusters and thereby the number of archetypes depends on their purpose: the ability to represent selected countries in the context of energy system followed by the application to an energy system model. Generating archetypes which are still explainable and manageable, while also reflecting the variety of input parameters for a reasonable application in an energy system model, is the overall challenge when determining the optimal number of clusters K .

2.1.3 Domain of validity – validation and benchmark process

The third step of archetype generation is defining the domain of validity. Our energy system archetypes are valid to describe countries by using data which characterizes their energy system. The archetypes are expected to be a better approach in summarizing countries regarding their energy system than a simple geographic classification.

To prove this hypothesis, a benchmark is performed after the clustering. The pure geographic classification in different studies is evaluated and compared to the clustering results. For this comparison, the average value of every clustered data category is calculated for each geographic region. Consequently, the Euclidean distance of every country within the geographic region to the average value is determined. By using the same distance measure as the clustering algorithm, this leads to a performance value which is comparable to the distances from the countries to the cluster centers of the

archetype. By comparing the distances of the clusters with the geographic classification, the plausibility of the clustering approach can be verified.

Since the archetypes are generated based on historical data, the energy system model anticipating a decarbonization transition for the future is not expected to be in the full domain of validity of archetypes which is determined by the clustering algorithm. For this reason, not all countries within the same archetype may well favor the same technologies in the future. Nevertheless, the correlation can be analyzed by comparing the model results of selected archetypes and its countries.

2.2 Automated model setup

The process of developing an automated model setup can be split in three sub-categories: general technological and economic assumptions, detailed modeling assumptions, and time series definition. The assumptions applied in the final modeling will be presented in the next sections of this chapter. More detailed information about the assumptions can be found in [34]. In general, the archetype models are generated by using the clustering results. Since many data categories cannot be compared in absolute values and are normalized for the clustering, a scaling process is necessary. The scaling is executed by using the mean population of all countries within an archetype. One big advantage of the presented method is that all other categories can be derived from this value. This relation makes it easy and fast to set up models, even for countries which do not provide much data. For the time series, lead countries are selected. These countries are the most similar countries to the archetype, indicated by the closest distance to the cluster centers.

2.2.1 General technological and economic assumptions

Technologies

As a first step, we focus on modeling the electricity sector which is decisive for the decarbonization due to the increasing electrification in other sectors [2]. Therefore, all storage technologies are re-converted to electricity. An overview of all technologies is provided in Figure 3.

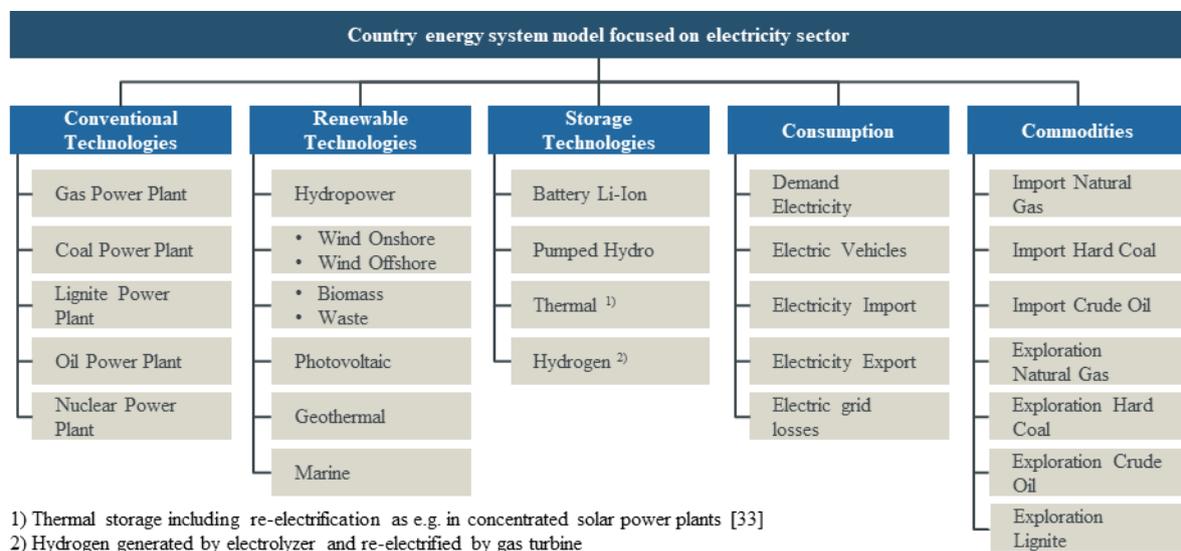


Figure 3. Overview setup of the energy system model

For the techno-economic analysis, the efficiency of the conversion processes and operation costs for the different technologies are taken from the Energy Technology Reference Indicator (ETRI) [35]. ETRI is chosen since it contains a broad range of technology parameters for many modeled technologies in one source which is necessary for a comparable model setup. If more than one power plant type is aggregated into one category, the average of them is taken. For some technologies, other sources are necessary, such as the information for nuclear, biomass, and the electrolyzer for hydrogen storage [36,37].

With regards to the commodities, they are either imported or extracted locally, as no exports are considered during the modelling process. In our approach, only one archetype is modeled at the same time. The trade of commodities could be further evaluated when modeling two archetypes in parallel and interconnecting these models. The electric grid losses are based on the current percentage they represent of the energy generation and maintained constant throughout the modeling time frame as no information of grid updates is available for each country. A reduction of the grid losses would affect mostly countries with a low GDP and reduce their energy demand by around 5-10% [38].

Fuel Prices

To facilitate the interpretation of results, fuel prices are assumed to be constant in time. They are mostly based on [39] while lignite prices are derived from [40]. Evaluating the specific costs of fossil fuels, the cheapest technology is lignite exploration, followed by hard coal exploration, hard coal import, uranium, gas exploration, gas import, oil exploration and oil import.

2.2.2 Detailed modeling assumptions

The development of a standardized approach requires to develop rules that take certain characteristics within an archetype and extend them to a general modeling parameter. The most relevant parameters will be discussed in this section. These can be divided in the base model assumptions, which are implemented for an accurate modeling of the present and future scenario. These are necessary to determine the transition pathways for decarbonization. Any of the following assumptions and model simplifications – just like any simplification – can be discussed. The set of simplifications presented here is the result of an intensive and iterative modelling effort on many energy systems worldwide.

Base modeling assumptions

Establishment of minimum full load hours: To model a realistic dispatch of power plants, considering outages, linkage to the heating sector, and transmission constraints, minimum full load hours are specified for gas, oil, and coal power plants. The minimum full load hours chosen for each archetype are defined by a global data analysis: based on historical data the generation of the mentioned technologies is calculated for each archetype and then assigned to its model.

Decommissioning of power plants: The aggregation of installed capacities into one technology category impedes the assignment of a construction year to a certain amount of capacity. To model the decommissioning of power plants, a linear approximation with constant slope is introduced. This slope is calculated individually for each technology by the installed capacity and the technical lifetime.

Future scenario definition

Electricity demand growth: In this modeling approach, economic and population growth determines the future energy demand of a country. The GDP for the target year is calculated by the growth rate per year which is a clustering result [34]. Using the correlation between GDP and energy intensity in [41], it is possible to determine the energy intensity per capita for 2050 by equation (1). When determining the absolute values, the projected population is used to determine the total electricity consumption for the country or archetype. Applying this formula, the archetypes differ in their projected GDP, but the factors a , b , and c are constant.

$$E_{cap}^{2050} = a \cdot \left(e^{b \cdot GDP_{cap}^{2050}} - e^{c \cdot GDP_{cap}^{2050}} \right) \quad (1)$$

$$a = 7.721 \cdot 10^4 [kWh], b = -1.95 \cdot 10^{-6} \left[\frac{1}{EUR} \right], c = -5.655 \cdot 10^{-6} \left[\frac{1}{EUR} \right]$$

To determine a reasonable development path, and to avoid the installation of resources where it is not technically feasible, the expansion of some technologies is limited. For each technology, an overview is presented below.

Nuclear: Due to the decommissioning plans of many countries worldwide and to the experienced delay in building new capacity, we assume that no nuclear is installed in the future [17,42,43].

Wind onshore and offshore: Since the installation of wind resources depends on land and resource availability, it is necessary to determine the allowed growth rate per year. By analyzing the archetype with the highest wind penetration, it is possible to define a normalized maximum annual expansion limit for these technologies.

For wind onshore, the installed capacity per km² is determined for 2015. For 2050, the estimated capacities from [30,44,45] are examined and the maximum value is selected. Expressing the installed capacity in these two years per unit area yields two points in time. The maximum annual expansion is determined by interpolating between these two values.

For wind offshore a similar approach is used with the coastline extension as the limiting factor. The density in MW/km is calculated for 2015 and 2050 to determine the expansion rate per annum. Wind offshore is only considered in countries which have a coastline.

For both technologies, the country with the fastest expansion is taken as a reference. Since absolute values are necessary to run the model, the absolute capacity is determined by multiplying the obtained slopes with the area or coastline extension of the archetype and country respectively. Equation (2)-(4) summarize the described density approach.

$$Density\ Onshore_{year} = \frac{Cap\ Installed_{year}[MW]}{Area\ [km^2]} \quad (2)$$

$$Density\ Offshore_{year} = \frac{Cap\ Installed_{year}[MW]}{Coastline\ [km]},\ if\ Coastline > 0 \quad (3)$$

$$Annual\ Expansion\ (slope) = \frac{Density_{2050} - Density_{2015}}{2050 - 2015} \quad (4)$$

Photovoltaic: Wind onshore and offshore expansion is limited by the "potential" due to the consideration of the available geographical area and extension. For photovoltaics (PV), a similar approach is used to determine the annual limits. In this case, the load estimated for 2050 is considered, as well as the FLH calculated from the clustering results. The necessary capacity to meet the total demand by just PV is calculated and divided by the number of years of the study period. This way, if the optimization model chooses PV as the cheapest option to cover total demand, it is able to do so in a feasible amount per year. This assumption provides the model a possibility to cover all demand by photovoltaic if it represents the most economical technology. The annual limit prevents that all photovoltaic is build towards the end of the decarbonization pathway.

Hydropower: Analyzing the development of hydropower over the years and given that it is an established technology, new power plants can only be installed where hydropower currently exists. The allowed expansion per annum is of 6.5% of the capacity installed in 2015 when the installed capacity is higher than 500 MW. For countries that have less than 500 MW installed, the maximum allowed expansion is 10 MW per annum. Both limits are obtained by analyzing the growth of hydropower plants in recent years [46]. To account for the different plant types, the availability of hydropower is split in a simplified approach: half of the installed capacity within an archetype or country is considered constantly generating while the other 50% can be dispatched in a flexible way.

Biomass: Since biomass is not limited by the same factors as wind or PV, its expansion limit is determined by its current share in the energy mix.

Geothermal, Waste, and Marine: For these technologies, 10% of the installed capacity in 2015 can be installed every year. This share is derived based on historical expansion rates and then scaled by the current use of each technology.

Renewable expansion rate: Due to the significant role the current share of renewables has in the rate of expansion, a correction factor is used. This factor is applied to all technologies except PV and hydropower and is introduced to allow countries with lower share of renewables in 2015 to install more capacity per annum. Thereby, the model is able to calculate the necessary installed capacities per year to reach the decarbonization goal of 80%. In Table 1 the correction factors are defined depending on the share of renewables. The shares and their respective correction factors are derived by analyzing the relationship between the current share and the allowed expansions in several model iterations.

Share of renewables x in %	Correction Factor
$x < 10$	4
$10 < x < 30$	3
$30 < x < 50$	2
$x > 50$	1

Table 1. Correction factor for allowed renewable expansion

Refurbishment of power plants: To account for the cost reduction caused by retrofitting of hydropower, wind, and PV plants (e.g. the use of existing grid connections, barrages in case of hydropower or components) a retrofit category is introduced [47,48]. The maximum amount of installed capacity available to be installed must be equal or lower than the amount of decommissioned capacity of its corresponding technology.

2.2.3 Time series definition

Due to the sensitivity that energy models have to time series [49], it is necessary to obtain information on the renewable resources' availability and energy demand fluctuation. Since the applied model calculates an hourly dispatch, demand and generation must be matched in every considered hour. For this paper, the time series information for PV, wind onshore, wind offshore, and demand are obtained for every country individually. The time step selection is chosen to model the year in 3-hour steps to account for a detailed representation of the year but limiting the computational complexity.

Load time series

The load profiles drive the generation necessities throughout the year and determine the moment and level of peak demand. For most European countries, the load profile can be taken from the information available on ENTSO-E's transparency platform [50]. For other countries load profiles are directly available online, e.g. for Mexico [51]. As a third step, if the information is not available, a synthetic profile is created using the profile of a typical day, month, and week during summer and winter. In this context for United Arab Emirates, information from [52-54] is combined. For India, the profile is created based on information from [55]. For all countries, which are not represented in the ENTSO-E data and where no individual data could be found or reconstructed, the load profile is taken from [41]. Since it is assumed that countries clustered within a certain archetype have similar energy systems, the profile of the lead country is used for detailed country analysis if no detailed profiles are available.

Photovoltaic and wind time series

The fluctuating profiles of wind and PV are all generated by the same source (renewables ninja) [56] but based on different methodologies. For wind offshore, data is either directly available or aggregated to generate a time series for each country [56]. It is important to note that the data on renewables ninja is not generation information, but rather is created by using weather and specific technology data to generate synthetic profiles of the availability of PV and wind [56-60].

The aggregation of the PV and wind time series is based on five points within the country (see Figure 4) but separated in two different approaches to calculate the respective weighting factors. These are both calibrated by comparing the developed model to real values for different countries.

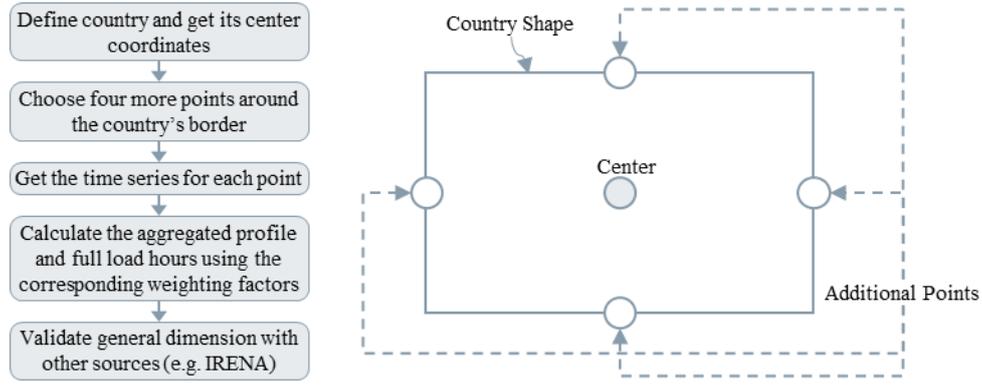


Figure 4. Process to determine wind and PV profile.

Wind onshore: To obtain the series for wind onshore, a turbine with 80m height as a typical size is considered. The weighting factors of the five points are assigned by the following rule: the center is given a 50% weight, while the other points are given a weight based on their annual capacity factors in the following order from best to worst location: 15%, 20%, 10%, and 5%. This approach is similar to [61], but the center is weighted more since it represents a larger area of the country.

PV: For this resource, a standard solar panel with direction to optimal infeed with no adjustable axes is considered. In this case, the center point does not have a higher weight, but the weighting factors are all assigned from best to worst location: 30%, 25%, 20%, 15%, 10%. This approach is selected since the differences between the inner value and outer boundaries are not as large as they can be for wind if a country is next to the coastline.

2.3 Performance evaluation

As described in Figure 1, the evaluation process consists of two steps: gathering the raw data and process it for evaluation purposes. The main output results are the installed capacity and dispatched energies for each technology, the electricity demand, CO₂ emissions, and the system costs. As a first processing step, the energy mix is calculated by dividing the generated amount of energy for each technology by the overall generated electricity. Storage technologies are normalized by the overall demand. By this normalization, archetypes and countries can be compared independently from their absolute demand. Furthermore, CO₂ emissions are normalized by the total electricity demand in order to find a comparable unit for this parameter. These two values, the energy mix and the normalized CO₂ emissions are primarily used for the comparison of archetypes. The results for all archetypes are compared and patterns are derived by analyzing the role of specific technologies.

As a second step, to evaluate the performance of the archetype model, evaluation criteria must be defined to quantify the level of accuracy. Therefore, the results of one chosen archetype model are compared to the results of each country within this archetype. The error metrics are applied on the energy mix of the country considering generation and storage. CO₂ emissions and costs are not considered since they directly depend on the energy mix and installed capacities.

In general, there are various performance measures to evaluate errors of models [62]. As a first performance metric, the mean absolute error (MAE) is chosen, which is in our case also a mean absolute percentage error. Both error categories are used very frequently [62]. The MAE is very easy to interpret since it directly evaluates the mean error of all evaluated technologies n . The calculation of the MAE is defined by equation (5). The variables A_j and P_j represent the actual and predicted values, which are in our case defined by the country value and the archetype value of technology j .

$$MAE = \frac{1}{n} \sum_{j=1}^n |A_j - P_j| \quad (5)$$

Another used metric is the root mean square error. This error metric is frequently used due to its relation to the standard deviation and large errors dominate the result [63]. However, when using this metric in

our current application, it allows for unimportant technologies with little share in the current energy mix to be disregarded. The calculation of this error type is given by equation (6):

$$RMSE = \sqrt{\frac{\sum_{j=1}^n (A_j - P_j)^2}{n}} \quad (6)$$

The two presented error metrics are calculated for each country within the evaluated archetype. The overall mean error of this archetype is defined by the average of all single errors of the countries it encompasses.

3. Results

As described in the framework (see Figure 2), the results occur in two parts: the archetypes as a result of the clustering process and the main results of applying these archetypes to an energy system model. This section contains both of these results separated in the following two sections 3.1 and 3.2.

3.1 Archetypes

In a first step, the collected data of more than 300 data categories is reduced by the data selection criteria. Table 2 provides an overview of the 62 categories which are considered for the clustering. Overall, the categories represent socio-economic (~20%) and climatic / geographic data (~20%) but the dataset is still mainly comprised of energy-related data (~60%) as the goal is to apply the clustering results to an energy system model.

Major topic	Data categories	Source
Socio-economic (14)	GDP per capita, GDP growth	[38]
	Human Development Index	[64]
	<u>GDP distribution</u> : agriculture, industry; political stability index, population density, urban population, urban population growth, access to electricity, rural access to electricity, time required to get electricity	[65]
	Electric vehicles per capita	[66]
	Population growth projection	[67]
Climatic / geographic (11)	Hydropower capacity factor	[63]
	Agricultural area, forest area	[65]
	Coast / area ratio	[68]
	Latitude absolute	[69]
	Average temperature, average precipitation	[70]
	Heating degree days, cooling degree days	[71]
	Solar capacity factor, wind capacity factor	[72]
Energy-related (37)	<u>Primary Energy intensity</u> : consumption per capita, economic intensity, consumption fossil fuels, consumption renewables; <u>Electricity</u> : consumption per capita, share consumption / installed capacity, share net imports, distribution losses; <u>Shares installed capacity</u> : fossil fuels, nuclear, renewables, hydropower, solar, geothermal, tidal and wave, hydroelectricity pumped storage; CO ₂ economic intensity; <u>Shares resources production / consumption</u> : petroleum, hard coal, lignite, natural gas	[38]
	<u>Shares installed capacity</u> : wind onshore, wind offshore, biomass, waste	[38,73]
	<u>Shares installed capacity</u> : oil, hard coal, lignite, natural gas	[38,74]
	Total natural resource rents, CO ₂ emissions per capita	[65]
	Emissions change from 1990	[75]
<u>Shares installed capacity storage</u> : total, electrochemical, electromechanical, thermal, hydrogen	[76]	

Table 2. Data categories used for cluster algorithm

Since most of the data is available for 2015, it is selected as the base year. Among the 193 UN countries, all data points of the 62 categories are available for 141 countries so that these can be included in the clustering algorithm. Other countries, which are not directly clustered, can subsequently be assigned to the closest cluster by the shortest Euclidean distance to the centroids but they are not further considered in this evaluation. An overview of the closest archetypes of these countries can be found in Table B.1 .

The number of archetypes, corresponding to the number of clusters K in the K-means algorithm, is set to 15 [19]. Thus, 15 archetypes still guarantee a certain level of detail and differentiation while still having explainable differences. Figure 5 provides an overview of the clustering results and thereby visualizes the 15 generated energy system archetypes.

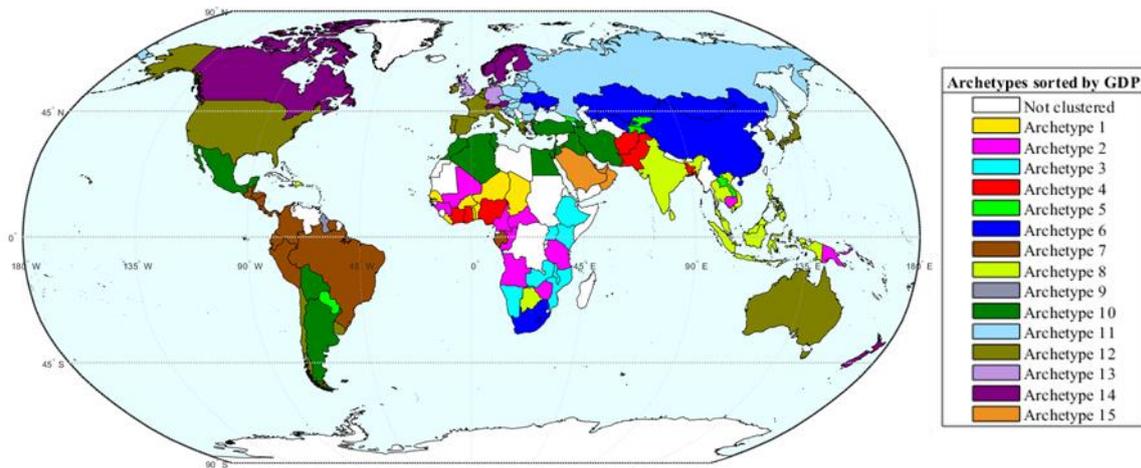


Figure 5. World map with 15 archetypes

Analyzing the distribution in the world map, the clustering leads to archetypes which are spread over different continents. This observation confirms the idea that clustering can summarize countries with a similar energy system, but which are not necessarily located in the same geographic region.

The detailed assignment of countries to the archetypes, including their lead countries and major characteristics, is listed in Table A.1 . In general, the idea of having easily explainable archetypes is met. For example, Archetype 5 summarizes countries which are all strongly dependent on hydropower and export a lot of electricity. Archetype 10 includes emerging countries whose energy system currently relies on fossil fuels but that has very good conditions for renewable energy penetration. Archetype 15 can be described as following: countries with relatively high energy and electricity consumption, high GDP, large oil and gas explorations (also used for electricity generation), and very good potential for solar power.

Next to the appendix, further details about the archetypes can be found in Figure 7 (a), showing the base results of the energy system model for 2015.

For validation, the generated and analyzed archetypes are then compared to the pure geographic regions which can e.g. be found in other energy system studies or are given by the UN [13-15,17,18,77]. Only the 141 countries which are considered in the clustering are evaluated in this context to ensure comparability. Figure 6 visualizes this comparison. The results prove that the clustering approach summarizes countries better when considering energy-related and socio-economic data in addition to the geographic analysis. In average, the improvement by clustering compared to a geographic classification is 44%.

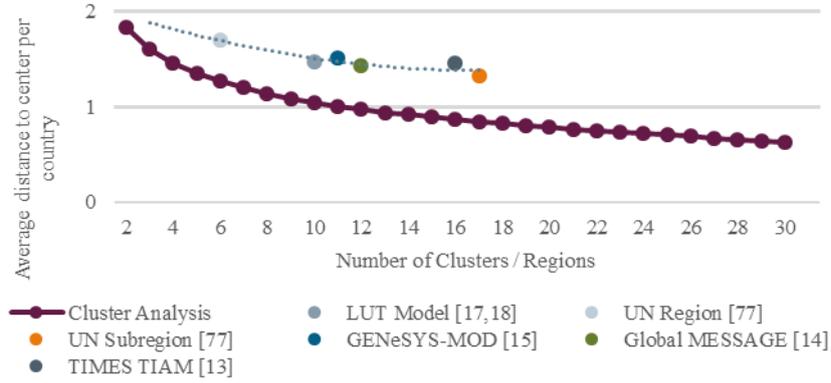


Figure 6. Comparison of cluster analysis with regional split

3.2 Modeling results analysis

The energy system archetypes are modeled by applying the modeling rules described in section 2.2. The modeling is done for a period spanning from 2015 (base year for clustering) to 2050 under an 80% decarbonization scenario. The resulting analysis follows the evaluation process described in section 2.3.

In a first step, the model assumptions which were developed for the automated model setup must be validated by comparing the results for 2015 with the real generation during the same year. This comparison is described in detail for Denmark, Germany and United Kingdom in [34]. The results show that overall the model represents the energy mix with an acceptable error. However, the dispatch of conventional power plans is sometimes not accurate although minimum full load hours were introduced.

3.2.1 Comparison of archetypes

After validating the general model assumptions through historical data, the model results for the decarbonization of the different archetypes can be analyzed in more detail. First, the general energy mix is shown in Figure 7 for the two years 2015 and 2050.

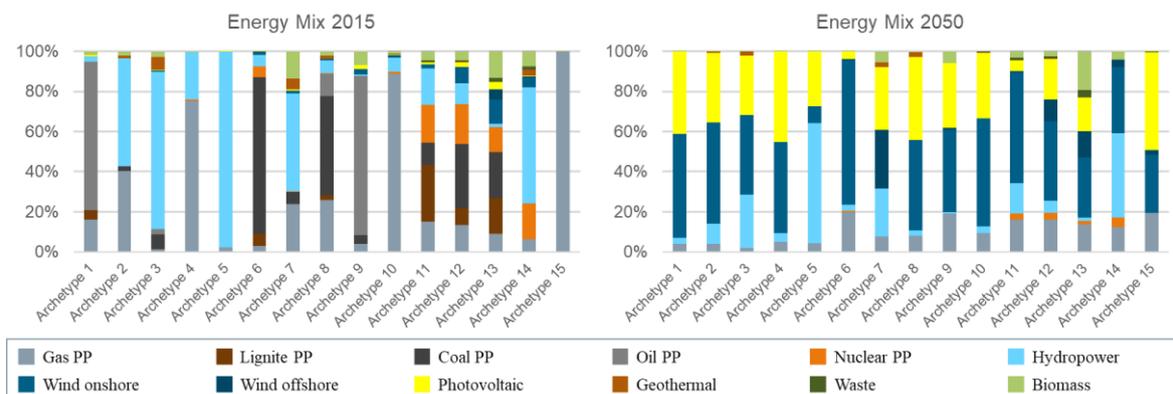


Figure 7. Generated energy mix of model results for 2015 (a) and 2050 (b)

As mentioned in the archetype explanation, the energy mix for the modeling results in 2015 (see Figure 7 (a)) reflects the different characteristics of all archetypes. The general comparison of the 15 archetypes indicates that the energy mix is totally different, even though social economic and climatic / geographic parameters also influenced the clustering. Analyzing exemplary archetypes in detail, this difference is confirmed: the oil and gas dependent countries in Archetype 15 only use gas power plants for electricity generation. Archetype 14, containing more Northern countries and New Zealand, has a large share of hydropower supplemented by nuclear power plants, and Archetype 5 is almost only using hydropower plants. Archetype 13, which represents Denmark, Germany, and the United Kingdom, has the highest share of wind onshore energy, while also using several conventional power plant technologies. Lastly,

Archetype 8 is an example of an archetype that does not primarily consist of industrialized countries. The electricity generation of these countries is dominated by fossil fuels as gas and coal power plants.

In comparison, Figure 7 (b) shows the development of the energy mix in an 80% decarbonization scenario until 2050 for each archetype. The decarbonization target leads to a limitation of available technologies which affects the final energy mix of the archetypes. For conventional generation, gas power plants are the only chosen option. Out of all available renewable technologies, hydropower, wind and PV play a major role for the energy systems. In the case of hydropower, the share of some archetypes becomes lower in 2050 than in 2015. This development is caused by an increase of the overall demand, especially in developing countries, which cannot be fully covered by new hydropower plants. As an example, the demand of Archetype 3 is eight times higher in 2050 than in 2015.

Even though at first glance the energy mixes look very similar in 2050, detailed evaluations indicate differences between the archetypes. Several interesting causalities were identified in the overall evaluation process of the model results. Three selected topics are presented in the following paragraphs to explain observed differences: (1) the combination of PV, hydropower and batteries, (2) the role of gas power plants and (3) sources for system flexibility.

Photovoltaics, hydropower and batteries

The first topic analyzes the share of PV in different archetypes and how the fluctuating generation profile of PV is balanced in the different systems. In this context, lithium-ion batteries and hydropower are identified as interesting technologies, whose share in the energy mix is shown in Figure 8 for all archetypes.

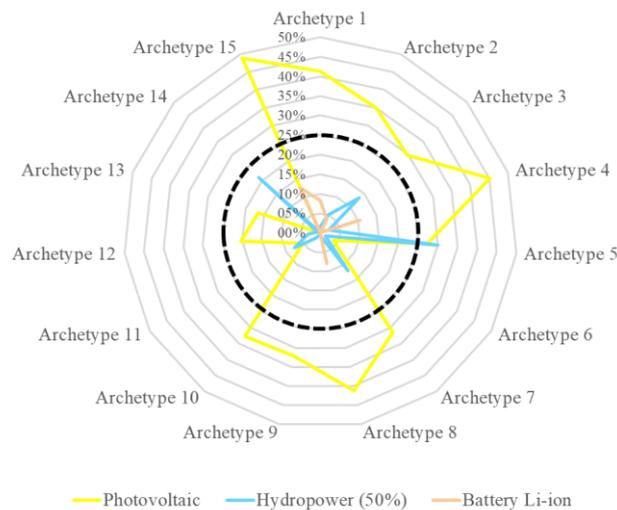


Figure 8. Result evaluation for photovoltaics, hydropower and batteries

In general, PV is an important technology for almost all archetypes. Only countries, whose geographical position is further away from the equator have a very low share of PV generation (Archetype 6, 11, 12, 13 and 14). For all mentioned archetypes the centroid of the absolute latitude is higher than 40° . To analyze the role of balancing technologies we exclude these five and focus on archetypes which have a share of PV generation higher than 25%, indicated by the dotted line in the diagram.

As a first conclusion, almost all archetypes which exceed this limit of 25% PV require a balancing technology for which lithium-ion batteries are suitable. Since costs for electrical power, and not stored energy, are low compared to those for other storage technologies [35], it is fit for daily cycles balancing an afternoon peak of PV to consumption in the evening or night. The same behavior can be seen in other analyses [36]. Nevertheless, batteries are not chosen for all archetypes with a share of PV generation higher than 25%. Archetypes with a noticeable share of flexible hydropower generation ($> 5\text{-}10\%$) do not require additional battery storage. These observations are confirmed by the comparison in Figure 9

analyzing the technology behavior in exemplary hours. Archetype 3, which has a high share of hydropower, uses this flexible share to balance the demand in the evening hours. In contrast, Archetype 15 balances the load by batteries which are charged during PV peaks.

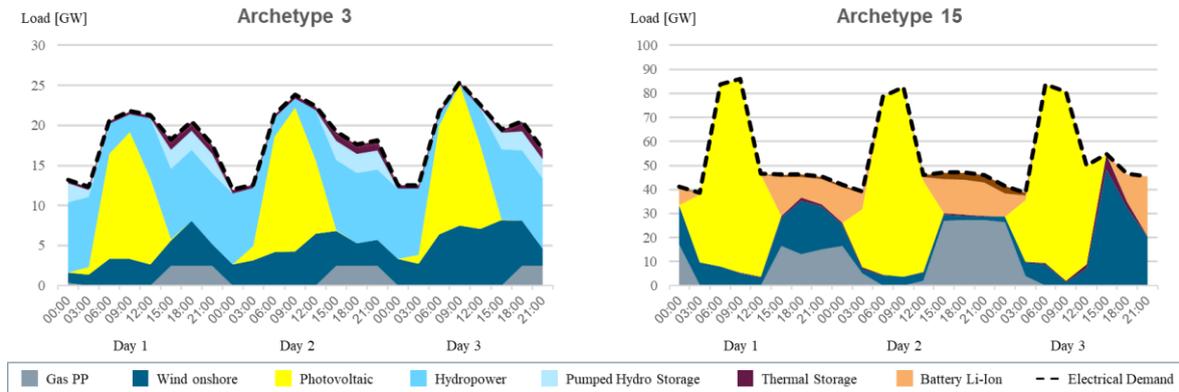


Figure 9. Comparison of three exemplary days for Archetype 3 (a) and 15 (b)

The only exception for the detected correlation of PV, batteries and hydropower is Archetype 9, which consists of several island states. There are mainly two reasons for this exception: (1) in 2050 the countries still have a comparably high share of gas power plants and (2) the wind capacity factor is high with a less fluctuating generation profile than in other archetypes. Overall, batteries are an important technology to integrate high shares of PV generation, especially if there is no flexible hydropower to substitute PV generation during the evening hours.

The role of gas power plants

The second analyzed topic is the role of gas power plants. Since we modeled an 80% decarbonization pathway, there is still some room for conventional generation units such as gas, oil, lignite or hard coal power plants. Thereof, gas power plants are the only chosen technology next to some nuclear power plants. Gas power plants are chosen since the specific CO₂ emissions are lower compared to the other options [35]. The limiting factor for the use of gas power plants are the allowed CO₂ emissions, calculated based on the emissions in 2015 [34]. The relation between the CO₂ emission cap and the share of gas power plants in the energy mix is shown in Figure 10.

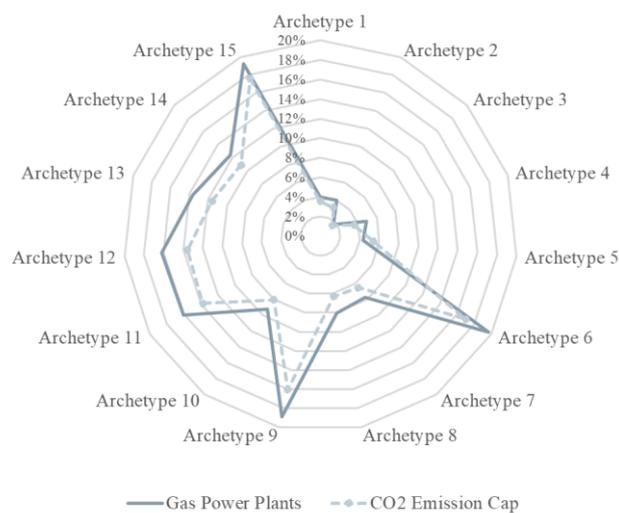


Figure 10. Result evaluation for gas power plants and CO₂ emissions

There are mainly two statements which can be derived from the analysis of gas power plants: first, all archetypes fully use the maximally allowed CO₂ emissions by generating electricity from gas. Secondly, the share of energy from gas power plants in the entire system and thereby the installed capacity

consequently depends on the emission cap. This conclusion can be clearly seen in Figure 10 as the two data series are closely correlated. As a result, archetypes which are currently emitting less also have a lower emission cap in 2050. Therefore, the share of gas power plants is also lower than in archetypes which are already emitting more (see Figure 10).

This comparison indicates that the approach of decarbonizing each archetype by 80% allows countries with currently high emissions, mostly industrialized countries, to emit a lot more per generated TWh electricity. On the other side, there might be a potential for developing countries to achieve a leapfrogging of certain technologies until 2050 and achieve a higher standard of electrification by lower emissions [78].

Sources for flexibility

As the share of fluctuating sources is high in all archetypes (see Figure 7), sources for flexible electricity generation and storage gain more importance. The most important sources, gas power plants, the flexible share of hydropower generation, and storage technologies are compared in Figure 11.

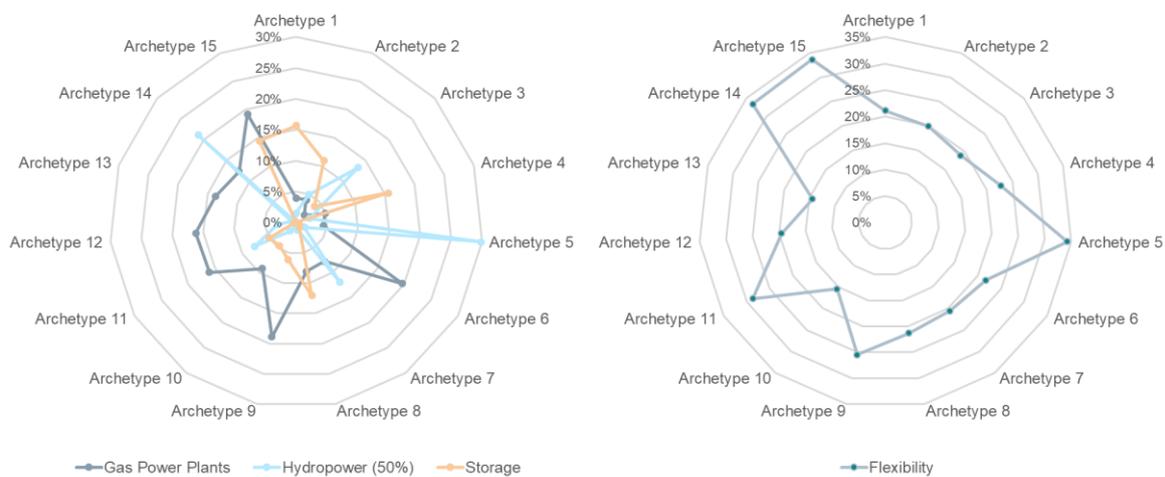


Figure 11. Result evaluation for flexibility sources (a) and summary of flexibility bandwidth (b)

The figure visualizes a higher need of flexibility ensured by storage technologies in archetypes with a low share of gas power plants, such as Archetypes 1, 2 and 4. In contrast, Archetype 3 compensates the need for flexibility by the flexible share of hydropower. The pattern of compensating lower gas shares with storage and hydropower can be applied to all other archetypes. Archetype 15 is an exception, as it has the second highest share of electricity generated from gas power plants but also from storage technologies. The reason for this exception has already been discussed before: the storage, in this case lithium-ion batteries, is needed to balance the daily peaks of PV generation to the load profile.

In Figure 11.b, the three analyzed technologies are summed up to an overall flexibility value. For most of the archetypes, this flexibility value is in a bandwidth of 15-22.5 percent – indicated by the colored corridor in the diagram. Countries which have a higher flexibility sum have a high share of hydropower generation. The two exceptions from this observation, Archetype 9 and 15, are already explained by analyzing PV and the overall flexibility. As a conclusion, all archetypes must ensure a certain level of flexibility to balance the higher share of fluctuating generation technologies. This minimum share is quantified to around 15% of the overall generation.

3.2.2 Deep dive into selected archetypes

In order to validate and better understand the archetypes, Archetype 8 and 15 are analyzed more detailed. The comparison is based on the modeling process depicted in Figure 2 and the performance evaluation described in section 2.3. As explained and visualized, the modeling of the countries is based on the same dataset used for clustering and the model is set up by applying the previously defined rules. The results of the performance evaluation for Archetype 15 are shown in Figure 12. In this figure, the energy mix

of the countries within this archetype and their errors compared to the archetype are shown. In the first column, the archetype model results are visualized in combination with the average error of the five countries.

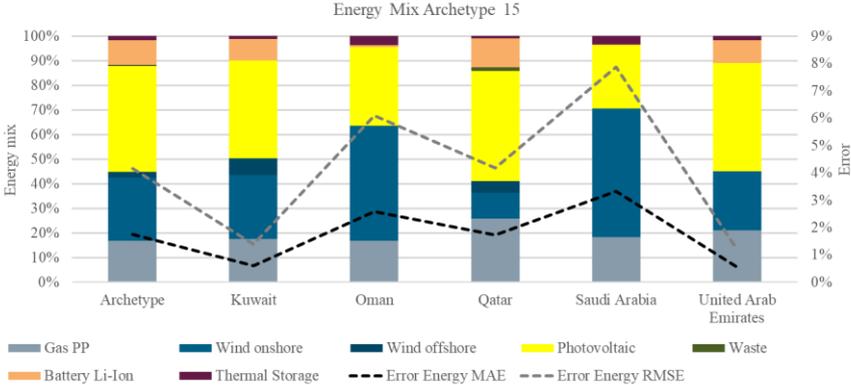


Figure 12. Comparison of archetype, its average error and country results for Archetype 15 in 2050

Overall, the comparison between the countries in this archetype proves that PV is an important technology. In all countries of this archetype the share of PV is higher than 25% which was identified as a threshold value in section 3.2.1. The medium average error for all considered technologies in the modeling process is 1.8% and the maximum 3.3%. The root mean squared error is between 1.2% and 7.9% with an average of 4.1%. The two countries which stand out in the error analysis are Oman and Saudi Arabia. There are two main reasons for these outliers: the wind generation profile and the overall demand. Concerning the wind profile, both countries have good wind conditions and especially a more balanced profile than the other three countries. While for these countries, the capacity factor is mostly in the range of 0-0.1 throughout the year, it is more equally distributed in Oman and Saudi Arabia. For the two countries the probability distribution of capacity factors has its peak values between 0.1 and 0.3. Additionally, the error gets higher for Saudi Arabia since it is by far the country with the highest demand in this archetype. In this case, a solution of PV and storage would be an economical challenge. As explained in section 3.2.1 the installation of PV requires a second balancing technology, which in this archetype are lithium-ion batteries.

As a second example, the performance evaluation of Archetype 8 for 2050 is shown in Figure 13. This archetype contains more countries leading to a higher variety of model results. In general, the higher number of countries leads to a higher average error. The medium average error is at 2.9% and the root mean squared error at 6.4%. The maxima rise respectively to 4.7% (MAE) and 10.7% (RMSE). Analyzing the major outliers characterized by the highest errors, four of the nine countries are focused: Botswana, Malaysia, Thailand and Vietnam.

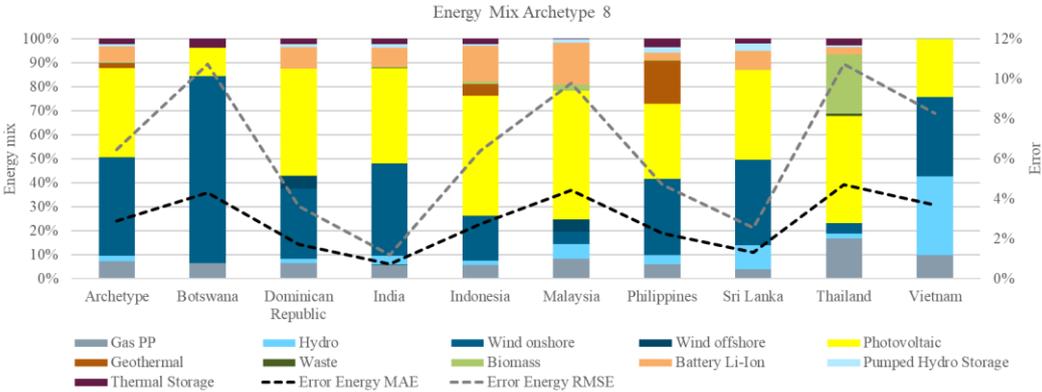


Figure 13. Comparison of archetype and country results for Archetype 8

In the case of Botswana, the demand is an important factor for the difference: since it is the lowest of all countries, a stable renewable system is needed to ensure a generation profile which meets the demand by avoiding a dominant share of fluctuating generation. A combination of PV and storage would lead to higher costs in this case. By contrast, the second and third outlier Malaysia and Thailand have worse wind conditions than the other countries within this archetype so that they favor other technologies: while Malaysia builds more PV and batteries, Thailand increases its share of electricity production by biomass. The last exception, Vietnam, has hydropower generation and therefore needs less PV as well as no battery storage to reach a decarbonization of 80%. This choice of technology also confirms the analyzed relation between PV, hydropower generation and batteries.

Overall a general trend of chosen technologies is recognizable for countries which belong to the same archetype. Nevertheless, the energy system model also leads to different results for some exceptions. These differences are expected, since the archetypes are generated based on historical data and on a broad range of data categories covering more aspects than the energy mix. However, the outliers are explainable by analyzing the country characteristics and the patterns derived from the model behavior.

The differences of countries within an archetype can be used to analyze sensitivities of the general approach and detect sensitive parameters for energy system models. These parameters are mainly the overall demand, the profiles, and the development of existing technologies. As a first parameter, the electricity demand influences the choice of technologies in both cases, comparably low and high demands. The examples of Saudi Arabia and Botswana proved this observation. Secondly, the profile is also important. Since not only the capacities but also the operation schedule is optimized, the demand needs to be covered even though shares of fluctuating generation sources are increasing. Therefore, technologies are sometimes favored if they have a balanced profile. This effect can be observed for the countries Oman and Saudi Arabia. A third observation is, that existing low-carbon technologies and their future potential play a major role, as explained by the examples of Thailand and Vietnam.

4. Conclusion

In this paper, an approach to model archetypes of country energy systems in an energy system optimization model was presented. Those models are frequently used to analyze medium- and long-term developments of country energy systems such as the decarbonization process. The presented approach of energy system archetypes simplifies analyzing the decarbonization process for worldwide countries by several advantages: it facilitates to model countries easily by a transferable database, enables data reconstruction from similar countries, provides an overview of countries facing similar challenges, and allows to evaluate the suitability of archetypal countries for specific technologies. As a basis, the archetypes were defined by a clustering approach and validated by applying error metrics as performance measures.

The classification of 141 countries into 15 archetypes summarizes countries by 44% better than a simple geographic classification in global energy system studies. As next to geographic characteristics, socio-economic and especially energy-related data play a major role for the energy systems, these three data categories were considered for the clustering and its evaluation.

The major results of modeling these archetypes firstly indicate that the energy mix of countries approximates in an 80% decarbonization scenario as the choice of technologies is more focused on renewable than on a mix of conventional and renewable generation technologies. However, the archetypes still develop differently, and a detailed analysis leads to three major results for future decarbonized energy systems: (1) If its share is higher than 25%, PV has a substantial need for balancing technologies regarding the daily cycle. This balancing is mostly ensured by batteries or hydropower. (2) Gas power plants are the only conventional generation technology but still chosen up to their allowed limit to provide a flexible generation source, and (3) flexibility is crucial as a bandwidth for flexibility providing technologies is determined (15-22.5%). Thereby, the archetype approach delivers a systematic analysis of future requirements and market size for technologies providing flexibility.

Lastly, the archetypes are validated by choosing two exemplary archetypes and comparing their model results to the respective countries. The results show a general match between the countries and the archetypes. As the clustering is based on historical data, outliers occur but can be explained by special country characteristics or the three mentioned major results. In general, the overall process leads to three sensitive parameters for the modeling process of future energy systems: (1) the overall electricity demand, (2) the profiles of demand and fluctuating energy sources, and (3) the development of existing carbon-free technologies.

This approach considered a global decarbonization transition by introducing energy system archetypes as a novel method. These archetypes could also help in world energy models by summarizing countries differently than just by their geographic region. Additionally, the classification in archetypes can be used for further analyses such as estimating market potentials. Further potential research is focused on two major aspects: firstly, the database and the modeling itself can be improved regarding financial aspects. There are already other studies which provide an overview how financing costs differ around the world [79,80]. Secondly, the modeling can be expanded by other sectors than the electricity sector. Since data is not easily accessible, methodologies to derive country-specific data are required.

Appendix A: Archetypes

Archetype	Countries (lead country highlighted)	Major characteristics (compared between archetypes)
Archetype 1	Benin, Burkina Faso, Chad, Comoros, Gambia , Liberia, Niger, Sao Tome and Principe, Senegal, Togo, Vanuatu	Lowest GDP/capita and HDI, highest average temperature, Lowest electricity consumption/capita and high dependency on fossil fuels
Archetype 2	Angola, Cambodia, Cameroon, Central African Republic, Republic of the Congo, Guinea, Mali, Papua New Guinea, Tanzania , Zimbabwe	Lowest population density, closest to equator, diverse energy mix (oil, gas, coal, hydropower)
Archetype 3	Ethiopia , Kenya, Malawi, Mozambique, Namibia, Uganda, Zambia	2 nd highest population growth, lowest urbanization, high share of hydropower, good solar conditions
Archetype 4	Afghanistan, Bangladesh, Cote d'Ivoire, Ghana, Nigeria, Pakistan	High population density, lowest political stability index, renewables only by hydropower (~30%)
Archetype 5	Albania , Georgia, Kyrgyzstan, Laos, Paraguay, Tajikistan	Low population density, Highest share of hydropower, highest share of electricity exports
Archetype 6	China, Kazakhstan , Moldova, Mongolia, South Africa, Ukraine, Uzbekistan	High share of hard coal and lignite, high energy intensity, mixed climate (heating / cooling)
Archetype 7	Belize, Brazil , Colombia, Costa Rica, Ecuador, El Salvador, Fiji, Gabon, Guatemala, Honduras, Nicaragua, Panama, Peru, Suriname	High precipitation, high shares of geothermal and biomass, coal exploration, high share of forests
Archetype 8	Botswana, Dominican Republic, India , Indonesia, Malaysia, Philippines, Sri Lanka, Thailand, Vietnam	High dependency on coal, large increase of CO2 emissions, low share of renewable generation technologies
Archetype 9	Bahamas, Barbados, Cape Verde, Guyana, Jamaica, Malta , Mauritius, Saint Kitts and Nevis, Samoa, Seychelles, Tonga	High population density, high impact of coastline (island states), high share of oil
Archetype 10	Algeria, Argentina, Azerbaijan, Bolivia, Cyprus, Egypt, Iran, Iraq, Israel, Jordan, Lebanon, Mexico , Morocco, Tunisia, Turkey	Low precipitation, high share of gas power plants, good combination of wind and solar capacity factors
Archetype 11	Armenia, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania , Russia, Slovakia, Slovenia	Shrinking population, low temperature, highest share of lignite, important role of nuclear
Archetype 12	Australia, Belgium, Chile, France, Greece, Ireland, Italy, Japan, Republic of Korea, Luxembourg, Netherlands, Portugal, Spain , United States, Uruguay	High urbanization, high shares of installed renewables, high emission intensity, medium solar and good wind conditions
Archetype 13	Denmark, Germany , United Kingdom	Highest HDI, highest shares of installed wind and PV, highest wind and lowest PV capacity factor, emission reduction since 1990
Archetype 14	Austria, Canada, Finland, New Zealand, Norway, Sweden , Switzerland	Lowest average temperature, highest share of electric vehicles, highest electricity consumption / capita, high share of primary energy from renewables, high share of hydropower and nuclear
Archetype 15	Kuwait, Oman, Qatar, Saudi Arabia, United Arab Emirates	Highest GDP / capita, highest energy consumption / capita, high temperature / low precipitation, highest share of fossil fuels (~100%), 2 nd best solar conditions, highest emission intensity

Table A.1 Archetypes and respective countries

Appendix B: Unclustered countries

Country	Closest Archetype
Antigua and Barbuda	Archetype 12
Bahrain	Archetype 15
Bhutan	Archetype 5
Brunei	Archetype 15
Burundi	Archetype 3
Cuba	Archetype 9
Democratic People's Republic of Korea	Archetype 6
Democratic Republic of the Congo	Archetype 3
Djibouti	Archetype 1
Dominica	Archetype 9
Equatorial Guinea	Archetype 8
Eritrea	Archetype 1
Grenada	Archetype 9
Guinea-Bissau	Archetype 1
Haiti	Archetype 1
Iceland	Archetype 14
Kiribati	Archetype 9
Lesotho	Archetype 3
Libya	Archetype 10
Macedonia	Archetype 11
Madagascar	Archetype 2
Maldives	Archetype 9
Mauritania	Archetype 1
Micronesia	Archetype 9
Montenegro	Archetype 5
Myanmar	Archetype 2
Nauru	Archetype 9
Nepal	Archetype 5
Rwanda	Archetype 3
Saint Lucia	Archetype 9
Saint Vincent and Grenadines	Archetype 9
Serbia	Archetype 11
Sierra Leone	Archetype 2
Singapore	Archetype 9
Solomon Islands	Archetype 1
Somalia	Archetype 1
South Sudan	Archetype 4
Sudan	Archetype 3
Swaziland	Archetype 4
Syria	Archetype 10
Timor-Leste	Archetype 2
Trinidad and Tobago	Archetype 15
Turkmenistan	Archetype 6
Tuvalu	Archetype 7
Venezuela	Archetype 7
Yemen	Archetype 4

Table B.1 Unclustered countries and corresponding archetypes

References

- [1] United Nations, 2015. Paris Agreement.
- [2] United Nations Environment Programme, 2019. Emissions Gap Report 2019. UNEP, Nairobi.
- [3] Lopian, P., Markewitz, P., Robinius, M., Stolten, D., 2018. A review of current challenges and trends in energy systems modeling. *Renewable and Sustainable Energy Reviews* 96, 156–166. <https://doi.org/10.1016/j.rser.2018.07.045>
- [4] Collins, S., Deane, J.P., Poncelet, K., Panos, E., Pietzcker, R.C., Delarue, E., et al., 2017. Integrating short term variations of the power system into integrated energy system models: A methodological review. *Renewable and Sustainable Energy Reviews* 76, 839–856. <https://doi.org/10.1016/j.rser.2017.03.090>
- [5] Després, J., Hadjsaid, N., Criqui, P., Noirot, I., 2015. Modelling the impacts of variable renewable sources on the power sector: Reconsidering the typology of energy modelling tools. *Energy* 80, 486–495. <https://doi.org/10.1016/j.energy.2014.12.005>
- [6] Mancarella, P., Andersson, G., Pecos-Lopes, J.A., Bell, K.R.W., 2016. Modelling of integrated multi-energy systems: Drivers, requirements, and opportunities. Presented at the 2016 Power Systems Computation Conference (PSCC), IEEE, Genoa, Italy, pp. 1–22. <https://doi.org/10.1109/PSCC.2016.7541031>
- [7] Sensuß, F., Genoese, M., Ragwitz, M., Möst, D., 2007. Agent-based Simulation of Electricity Markets -A Literature Review-. *Energy Studies Review* 15. <https://doi.org/10.15173/esr.v15i2.507>
- [8] Möst, D., Keles, D., 2010. A survey of stochastic modelling approaches for liberalised electricity markets. *European Journal of Operational Research* 207, 543–556. <https://doi.org/10.1016/j.ejor.2009.11.007>
- [9] Pfenninger, S., Hirth, L., Schlecht, I., Schmid, E., Wiese, F., Brown, T., et al., 2018. Opening the black box of energy modelling: Strategies and lessons learned. *Energy Strategy Reviews* 19, 63–71. <https://doi.org/10.1016/j.esr.2017.12.002>
- [10] Pfenninger, S., Hawkes, A., Keirstead, J., 2014. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews* 33, 74–86. <https://doi.org/10.1016/j.rser.2014.02.003>
- [11] Savvidis, G., Siala, K., Weissbart, C., Schmidt, L., Borggreffe, F., Kumar, S., et al., 2019. The gap between energy policy challenges and model capabilities. *Energy Policy* 125, 503–520. <https://doi.org/10.1016/j.enpol.2018.10.033>
- [12] Pfenninger, S., DeCarolis, J., Hirth, L., Quoilin, S., Staffell, I., 2017. The importance of open data and software: Is energy research lagging behind? *Energy Policy* 101, 211–215. <https://doi.org/10.1016/j.enpol.2016.11.046>
- [13] Vaillancourt, K., Labriet, M., Loulou, R., Waub, J.-P., 2008. The role of nuclear energy in long-term climate scenarios: An analysis with the World-TIMES model. *Energy Policy* 36, 2296–2307. <https://doi.org/10.1016/j.enpol.2008.01.015>
- [14] Rogner, H., 2016. IASA's Integrated Assessment Framework. Presented at the Pathways to Sustainable Energy Workshop. Geneva
- [15] Löffler, K., Hainsch, K., Burandt, T., Oei, P.-Y., Kemfert, C., von Hirschhausen, C., 2017. Designing a model for the global energy system—GENeSYS-MOD: An Application of the Open-Source Energy Modeling System (OSeMOSYS). *Energies* 10, 1468. <https://doi.org/10.3390/en10101468>
- [16] Aboumahboub, T., 2012. Modeling and Optimization of the Global Electricity Generation System with High Shares of Fluctuating Renewable Energy Sources (PhD). Technische Universität München, München.
- [17] Bogdanov, D., Farfan, J., Sadovskaia, K., Aghahosseini, A., Child, M., Gulagi, A., et al., 2019. Radical transformation pathway towards sustainable electricity via evolutionary steps. *Nat Commun* 10, 1077. <https://doi.org/10.1038/s41467-019-08855-1>
- [18] Ram, M., Bogdanov, D., Aghahosseini, A., Oyewo, S., Gulagi, A., Child, et al., 2017. Global Energy System based on 100% Renewable Energy - Power Sector. Lappeenranta University of Technology and Energy Watch Group, Lappeenranta; Berlin.
- [19] Küppers, M., Metzger, M., Huber, M., Paulus, S., 2019. Archetypes of Country Energy Systems, in: 2019 IEEE Milan PowerTech. Presented at the 2019 IEEE Milan PowerTech, IEEE, Milan, Italy, pp. 1–6. <https://doi.org/10.1109/PTC.2019.8810765>
- [20] Norwegian Ministry of Petroleum and Energy, 2020. Energy Facts Norway [WWW Document]. URL <https://energifaktanorge.no/en/>
- [21] Wogan, D., Pradhan, S., Albardi, S., 2017. GCC Energy System Overview – 2017.
- [22] Oberlack, C., Sietz, D., Bürgi Bonanomi, E., de Bremond, A., Dell'Angelo, J., Eisenack, K., et al., 2019. Archetype analysis in sustainability research: meanings, motivations, and evidence-based policy making. *E&S* 24, art26. <https://doi.org/10.5751/ES-10747-240226>
- [23] Atalla, T., Bean, P., 2017. Determinants of energy productivity in 39 countries: An empirical investigation. *Energy Economics* 62, 217–229. <https://doi.org/10.1016/j.eneco.2016.12.003>
- [24] Cserekllyei, Z., Thurner, P.W., Langer, J., Küchenhoff, H., 2017. Energy paths in the European Union: A model-based clustering approach. CCEP Working Paper 1701, Crawford School of Public Policy, The Australian National University.
- [25] Weinand, J.M., McKenna, R., Fichtner, W., 2019. Developing a municipality typology for modelling decentralised energy systems. *Utilities Policy* 57, 75–96. <https://doi.org/10.1016/j.jup.2019.02.003>
- [26] De Vivero, G., Burges, K., Kurdziel, M.-J., Hagemann, M., 2019. Transition towards a decarbonised electricity sector - A framework of analysis for power system transformation. NewClimate Institute.
- [27] Stuckenberger, P., Wildemann, H., Blomberg, C., Engelmeier, T., Hellmann, C., Offizier, M.A., 2013. Connecting Possibilities - Scenarios for Optimizing Energy Systems. Siemens Energy Sector, Erlangen.
- [28] Eisenack, K., Villamayor-Tomas, S., Epstein, G., Kimmich, C., Magliocca, N., Manuel-Navarrete, D., et al., 2019. Design and quality criteria for archetype analysis. *E&S* 24, art6. <https://doi.org/10.5751/ES-10855-240306>
- [29] Raths, S., Koopmann, S., Müller, C., Meinerzhagen, A., Falke, T., Cramer, M., et al., 2015. The Energy System Development Plan (ESDP). Presented at the ETG Kongress 2015, p. 8.
- [30] Müller, C., Hoffrichter, A., Wyrwoll, L., Schmitt, C., Trageser, M., Kulms, T., et al., 2019. Modeling framework for planning and operation of multi-modal energy systems in the case of Germany. *Applied Energy* 250, 1132–1146. <https://doi.org/10.1016/j.apenergy.2019.05.094>
- [31] Jain, A.K., 2010. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31, 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- [32] Patel, V.R., Mehta, R.G., 2011. Impact of outlier removal and normalization approach in modified k-Means clustering algorithm. *IJCSI International Journal of Computer Science Issues* 8.
- [33] Arthur, D., Vassilvitskii, S., 2007. k-means++: The Advantages of Careful Seeding, in: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, pp. 1027–1035.

- [34] Paredes Pineda, S.N., 2019. Development and Evaluation of a Simplified Approach for Modeling Country Energy Systems (Master Thesis). Technische Universität München, München.
- [35] Carlsson, J., n.d. ETRI 2014 - Energy Technology Reference Indicator projections for 2010-2050 (EUR - Scientific and Technical Research Reports No. JRC92496). DOI: [10.2790/057687](https://doi.org/10.2790/057687)
- [36] Walter, O., Huber, M., Kueppers, M., Tremel, A., Becker, S., 2019. Energy system design for deep decarbonization of a sunbelt city by using a hybrid storage approach, in: Proceedings of the 13th International Renewable Energy Storage Conference 2019 (IRES 2019). Presented at the Proceedings of the 13th International Renewable Energy Storage Conference 2019 (IRES 2019), Atlantis Press, Düsseldorf, Germany. <https://doi.org/10.2991/ires-19.2019.23>
- [37] European Climate Foundation, 2010. Roadmap 2050: a practical guide to a prosperous, low-carbon Europe. European Climate Foundation.
- [38] U.S. Energy Information Administration, International Energy Statistics [WWW Document]. URL <https://www.eia.gov/beta/international/data/browser> (accessed Nov. 18).
- [39] Vidal R., 2018. IHS long-term planning and energy scenarios – IHS Energy Autonomy.
- [40] Görner, K., Sauer, D.U., 2016. Konventionelle Kraftwerke: Technologiesteckbrief zur Analyse „Flexibilitätskonzepte für die Stromversorgung 2050“.
- [41] Toktarova, A., Gruber, L., Hlusiak, M., Bogdanov, D., Breyer, C., 2019. Long term load projection in high resolution for all countries globally. International Journal of Electrical Power & Energy Systems 111, 160–181. <https://doi.org/10.1016/j.ijepes.2019.03.055>
- [42] Child, M., Koskinen, O., Linnanen, L., Breyer, C., 2018. Sustainability guardrails for energy scenarios of the global energy transition. Renewable and Sustainable Energy Reviews 91, 321–334. <https://doi.org/10.1016/j.rser.2018.03.079>
- [43] Sovacool, B.K., Gilbert, A., Nugent, D., 2014. Risk, innovation, electricity infrastructure and construction cost overruns: Testing six hypotheses. Energy 74, 906–917. <https://doi.org/10.1016/j.energy.2014.07.070>
- [44] Barnacle, M., Robertson, E., Galloway, S., Barton, J., Ault, G., 2013. Modelling generation and infrastructure requirements for transition pathways. Energy Policy 52, 60–75. <https://doi.org/10.1016/j.enpol.2012.04.031>
- [45] Buhl, J.H., Hegelund, A.Z., Simonsen, M.B., 2017. Roadmap to 2050: A recommended future Danish energy system and guidelines for policy changes (Master Thesis). University of Southern Denmark.
- [46] International Hydropower Association (IHA), 2019. Hydropower status report: sector trends and insights.
- [47] Van Vuuren, S., Bliersch, C., Van Dijk, M., 2011. Modelling the feasibility of retrofitting hydropower to existing South African dams. WSA 37, 679–692. <https://doi.org/10.4314/wsa.v37i5.5>
- [48] Kougiass, I., Aggidis, G., Avellan, F., Deniz, S., Lundin, U., Moro, A., et al., 2019. Analysis of emerging technologies in the hydropower sector. Renewable and Sustainable Energy Reviews 113, 109257. <https://doi.org/10.1016/j.rser.2019.109257>
- [49] Staffell, I., Pfenninger, S., 2018. The increasing impact of weather on electricity supply and demand. Energy 145, 65–78. <https://doi.org/10.1016/j.energy.2017.12.051>
- [50] ENTSO-E, 2019. ENTSO-E Transparency Platform [WWW Document]. URL <https://transparency.entsoe.eu/dashboard/show>
- [51] Gobierno De Mexico, 2019. Estimación de la Demanda Real del Sistema [WWW Document]. URL <https://www.cenace.gob.mx/SIM/VISTA/REPORTES/DemandaRealSist.aspx>
- [52] Smith, B., 2018. The Role of Energy Storage in Planning our Energy Needs in the UAE.
- [53] Gastli, A., Charabi, Y., Alammari, R.A., Al-Ali, A.M., 2013. Correlation between climate data and maximum electricity demand in Qatar, in: 2013 7th IEEE GCC Conference and Exhibition (GCC). Presented at the 2013 7th IEEE GCC Conference and Exhibition (GCC), IEEE, Doha, Qatar, pp. 565–570. <https://doi.org/10.1109/IEEEGCC.2013.6705841>
- [54] Gregorio, A.D., Zahr, A.A., 2015. Dubai Demand Side Management Strategy: 2015 Annual Report “The future is efficient”
- [55] Soonee, S.K., Baba, K.V.S., Narasimhan, S.R., Nallarasan, N., Reddy, M.P., Chilukuri, P., 2016. Flexibility Requirement in Indian Power System
- [56] Staffell, I., Pfenninger, S., 2016. Using bias-corrected reanalysis to simulate current and future wind power output. Energy 114, 1224–1239. <https://doi.org/10.1016/j.energy.2016.08.068>
- [57] Müller, R., Pfeifroth, U., Träger-Chattejee, C., Trentmann, J., Cremer, R., 2015. Digging the METEOSAT Treasure—3 Decades of Solar Surface Radiation. Remote Sensing 7, 8067–8101. <https://doi.org/10.3390/rs70608067>
- [58] The Climate Data Guide: Surface Solar Radiation Data Set - Heliosat (SARAH) -Edition 1 [WWW Document], n.d. URL <https://climatedataguide.ucar.edu/climate-data/surface-solar-radiation-data-set-heliosat-sarah-edition-1>
- [59] Pfenninger, S., Staffell, I., 2016. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy 114, 1251–1265. <https://doi.org/10.1016/j.energy.2016.08.060>
- [60] Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., et al., 2011. MERRA: NASA’s Modern-Era Retrospective Analysis for Research and Applications. J. Climate 24, 3624–3648. <https://doi.org/10.1175/JCLI-D-11-00015.1>
- [61] Breyer, C., 2017. 100% Renewables Scenarios: Model, Data and Results. <https://doi.org/10.13140/RG.2.2.18118.88641>
- [62] Botchkarev, A., 2019. A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. IJIKM 14, 045–076. <https://doi.org/10.28945/4184>
- [63] Li, X., Zhao, Z., 2006. Evaluation of estimation algorithms part I: incomprehensive measures of performance. IEEE Trans. Aerosp. Electron. Syst. 42, 1340–1358. <https://doi.org/10.1109/TAES.2006.314576>
- [64] United Nations Development Programm, Human Development Reports [WWW Document]. URL <http://hdr.undp.org/en/data> (accessed Nov. 18).
- [65] The World Bank, World Development Indicators [WWW Document]. URL <https://databank.worldbank.org/source/world-development-indicators> (accessed Nov. 18).
- [66] wattEV2Buy, Global EV Sales [WWW Document]. URL <https://wattEV2buy.com/global-ev-sales/> (accessed Nov. 18).
- [67] United Nations Department of Economic and Social Affairs, World Population Prospects 2019 [WWW Document]. URL <https://population.un.org/wpp/> (accessed 3.1.19).
- [68] Gscociology, International data sets [WWW Document]. URL <http://gsociology.icaap.org/dataupload.html> (accessed Nov. 18).
- [69] Dataset Publishing Language, Countries.csv [WWW Document]. URL https://developers.google.com/public-data/docs/canonical/countries_csv (accessed Nov.18).
- [70] World Bank Group, Climate Change Knowledge Portal [WWW Document]. URL <https://climateknowledgeportal.worldbank.org/> (accessed Nov. 18).
- [71] Baumert, K., Selman, M., 2003. Heating and Cooling Degree Days [WWW Document]. URL <https://www.scribd.com/document/34318847/Degree-Days> (accessed Nov. 18).

- [72] Pfenninger, S., Staffell, I., Renewables ninja [WWW Document]. URL <https://www.renewables.ninja/> (accessed Mar. 19).
- [73] International Renewable Energy Agency (IRENA), Installed Renewable Power Capacity [WWW Document]. URL <http://dashboard.irena.org/> (accessed Mar. 19).
- [74] World Resources Institute, Global Power Plant Database [WWW Document]. URL <http://datasets.wri.org/dataset/globalpowerplantdatabase> (accessed Jul. 19).
- [75] World Resources Institute, CAIT Climate Data Explorer [WWW Document]. URL <http://cait.wri.org/> (accessed Nov. 18).
- [76] Office of Electricity Delivery & Energy Reliability, DOE Global Energy Storage Database [WWW Document]. URL <https://www.energystorageexchange.org/projects> (accessed Nov. 18).
- [77] United Nations, Geographic Regions [WWW Document]. URL <https://unstats.un.org/unsd/methodology/m49/> (accessed Jul. 19).
- [78] Szabó, S., Bódis, K., Huld, T., Moner-Girona, M., 2013. Sustainable energy planning: Leapfrogging the energy poverty gap in Africa. *Renewable and Sustainable Energy Reviews* 28, 500–509. <https://doi.org/10.1016/j.rser.2013.08.044>
- [79] Steffen, B., 2019. Estimating the Cost of Capital for Renewable Energy Projects. *SSRN Journal*. <https://doi.org/10.2139/ssrn.3373905>.
- [80] Ondraczek, J., Komendantova, N., Patt, A., 2015. WACC the dog: The effect of financing costs on the levelized cost of solar PV power. *Renewable Energy* 75, 888–898. <https://doi.org/10.1016/j.renene.2014.10.053>.