ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Kittel, Martin; Hobbie, Hannes; Dierstein, Constantin

Article — Accepted Manuscript (Postprint)

Temporal aggregation of time series to identify typical hourly electricity system states: A systematic assessment of relevant cluster algorithms

Energy

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Kittel, Martin; Hobbie, Hannes; Dierstein, Constantin (2022) : Temporal aggregation of time series to identify typical hourly electricity system states: A systematic assessment of relevant cluster algorithms, Energy, ISSN 1873-6785, Elsevier, Amsterdam, Vol. 247, pp. 1-15,
https://doi.org/10.1016/j.energy.2022.123458

This Version is available at:
https://hdl.handle.net/10419/284363

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.


https://creativecommons.org/licenses/by-nc-nd/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

This is the postprint of an article published in: Energy 247 (2022), 123458, 15 S.
Available online at: https://doi.org/10.1016/j.energy.2022.123458

Temporal Aggregation of Time Series to Identify Typical Hourly Electricity System States: A Systematic Assessment of Relevant Cluster Algorithms

Martin Kittel[a,*], Hannes Hobbie[b], Constantin Dierstein[b]

[a] *DIW Berlin, Department of Energy, Transportation, Environment, Mohrenstraße 58, 10117 Berlin, Germany*
[b] *TU Dresden, Chair of Energy Economics, Münchner Platz 3, 01069 Dresden, Germany*

Abstract

Comprehensive numerical models are pivotal to analyze the decarbonization of electricity systems. However, increasing system complexity and limited computational resources impose restrictions to model-based analyses. One way to reduce computational burden is to use a minimum, yet representative, set of system states for model simulation. These states characterize fluctuating renewable generation and variable demand for electricity prevailing at a certain point in time. A review of possible time series aggregation techniques identifies cluster algorithms as most adequate, with k-Means and the Ward algorithm predominating. However, throughout the surveyed literature, the line of reasoning for the selection of these algorithms remains unclear. To support the electricity system modeling community in selecting an algorithm, this paper devises a systematic multi-stage evaluation approach to compare a large variety of cluster analysis configurations, differing in algorithm, cluster representation and number of clusters. Results show that electricity demand and renewable energy generation time series can be compressed to below one percent while sustaining global characteristics of the original data. Two potent cluster configurations are identified, confirming k-Means and Ward as being prevalent. Beyond electricity market data, the methodology can be applied to various types of fundamental time-dependent input data.

Keywords: Cluster Analysis, Time Series Aggregation, Variable Renewable Energy, Electricity Market Modeling, Typical System States
2010 MSC: 62H30

1. Introduction

Electricity systems are undergoing major structural changes due to the rising penetration of renewable energy, additional demand-side flexibilities, or the electrification of the transport and heating sector. Electricity market models are used to investigate these developments. They become increasingly complex, entailing additional computational burden. A comprehensive representation might require a reduction of

*Corresponding author
Email addresses: mkittel@diw.de (Martin Kittel), hannes.hobbie@tu-dresden.de (Hannes Hobbie), constantin.dierstein@tu-dresden.de (Constantin Dierstein)

© <2022>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

model size to handle complexity. Operational electricity system states[1] refer to the prevailing average variable renewable energy sources (VRE) penetration and electricity demand levels of one specific time period, mostly one hour. To some extent, they are repetitive over the course of one year. A temporal aggregation of similar or identical system states by identifying a set of representative system states may reduce model size. Additionally, these system states potentially provide valuable insights into market structures.

Aggregating the temporal resolution of VRE generation and load profiles requires a systematic selection process. Choosing an insufficient set of unrepresentative system states may lead to an underestimation of VRE variability and overestimated VRE availability. Overrating the role of VRE results in biased model results [2].

One common technique for temporal aggregation of fundamental input data in electricity market modeling is cluster analysis. Two cluster algorithms are primarily used: Ward's Minimum Variance Method (WARD) algorithm from the class of hierarchical cluster algorithm (HACAL) and k-Means from the class of partitional cluster algorithm (PARTAL). The objective rationale of the cluster algorithms itself could motivate their application. However, relevant studies applying these algorithms lack a thorough explanation as to why HACAL or PARTAL was chosen as preferred cluster class, and, furthermore, why WARD or k-Means was selected over other available algorithms within the same class. Consequently, the line of reasoning for the selection of WARD or k-Means for temporal aggregation in electricity market modeling remains unclear. This paper aims at closing this gap in the literature. The overarching methodological objective is to systematically investigate HACAL and PARTAL with respect to representation quality of the original input data. To this end, a large number of different cluster scenarios is scrutinized, each differing in terms of cluster algorithm, cluster representation, and the optimal number of clusters within a cluster solution. Assessing the obtained clusters in terms of representation quality of the original input data, evaluation criteria allow for meaningful interpretation of a cluster solution. Traditionally, this is based on statistical key indicators. Although evaluating to what extent the approximation reflects global statistic properties of the original data, they cannot reveal whether clustered time series are performant in application. This paper employs a three-stage fundamental data clustering and electricity market modeling approach. It complements a statistical evaluation of cluster algorithms with an assessment of their model-based performance. For the latter, a flow-based electricity market model for a case study of the Central West European (CWE) region is used.

A major merit of the combination of statistical key indicators and application-based result information is that it allows for an extensive investigation. 260 initial cluster scenarios are narrowed down to the 100 most promising ones. Subsequently, two salient cluster scenarios from the HACAL and PARTAL class are identified. Each generates a minimum, yet representative, set of hourly system states for system representation.

[1] In literature, alternative expressions for system states are commonly used (e.g. snap shot, or time-slice), which have different period lengths (e.g. a complete week or day, multiple consecutive hours, or hourly system states) [1].

The remainder of the paper is structured as follows: Section 2 outlines empirical findings for temporal aggregation methods used in electricity market modeling. Section 3 details the design and subject of this research. Section 4 discusses statistical characteristics of obtained cluster solutions. Section 5 investigates their application-based performance. Section 6 analyzes the determined hourly system states. Finally, Section 7 summarizes results and discusses implications for electricity market modeling. Additional material is available in the supplementary information (SI), such as a brief introduction into the theory on cluster analysis.

2. Empirical findings and research focus

This section elaborates on temporal and spatial variability in electricity market modeling and related empirical findings. Based on this, the section specifies the research focus of this paper.

2.1. Temporal and spatial variability in electricity market modeling

Crucial aspects when aggregating intra-annual temporal resolution are spatial and temporal variability within the investigated period. Spatial variability must be taken into account when analyzing multi-regional settings. Traditionally, the major driver of temporal variability in electricity markets was fluctuations in electricity demand, e.g. variations between day- and night-time or between seasons. The introduction of VRE, such as wind or solar photovoltaic (PV), induced further variability into the system. The role of VRE is set to increase in future, not only in the electricity sector but also for decarbonizing mobility, heating, and industrial processes [3]. The adequate treatment of variability gains significantly in importance in electricity market modeling.

While there are studies that do not aggregate at all (e.g. [4]), the aggregation of temporal resolution for electricity market modeling is recognized in numerous studies employing different aggregation approaches. Their underlying rationale is to reduce the temporal resolution of an original reference data set, while maintaining the global properties of the reference data, such as temporal variability, spatial variability, and concomitant extremes. This paper focuses on the aggregation of demand and VRE generation patterns for electricity market modeling. Figure 1[2] illustrates three major approaches to derive hourly system states [5]: Random selection, heuristic selection, and aggregation based on cluster analyses, with ascending methodological complexity.

The idea of a random selection is to derive a subset of hourly system states based on arbitrary or random pick. It comes with very low implementation requirements, yet, is highly affected by the number and selection of system states.

Heuristic approaches use more advanced selection procedures to determine representative system states. Content related systematics determine system states based on the fundamental features of the reference data set, such as seasonality within the year, week, or day. Hourly whereas system states can

[2] A selection of relevant literature is presented. Related literature on temporal aggregation for modeling building energy systems is comprehensive, yet, beyond the focus of this paper.

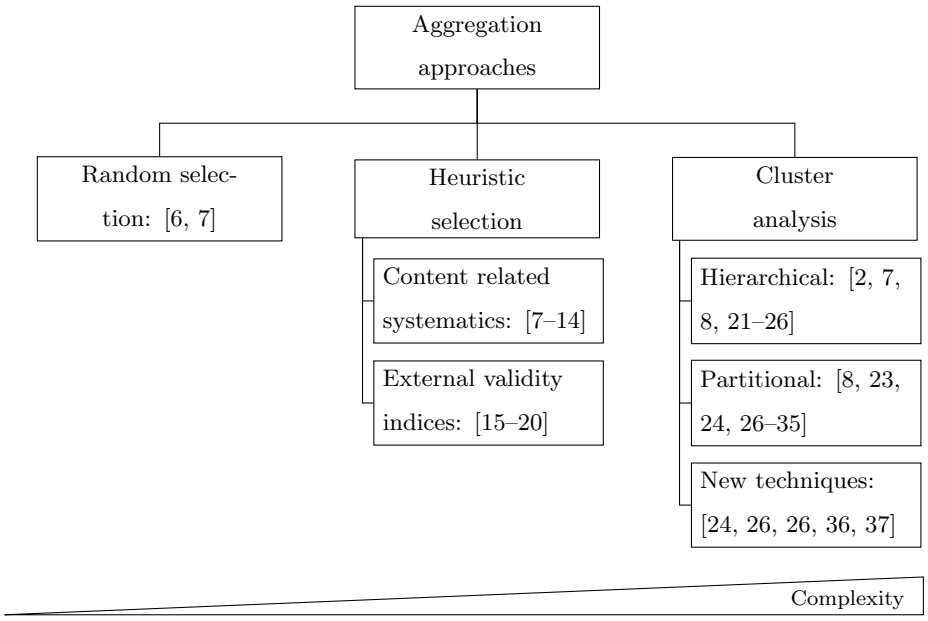


Figure 1: Classification of time resolution aggregation approaches including assigned research.

also be selected based on external validity indices, e.g. statistical error metrics. The applicability of heuristics is limited. The selection of a suitable heuristic strongly hinges upon the user's depth and understanding of domain knowledge, thus is susceptible to subjective bias.

Finally, cluster methods aim at identifying structures in a reference data set by objectively allocating all observations into homogeneous groups called clusters. From each cluster, a representative is selected and used as an input system state for an electricity market modeling. Since cluster analyses follow a predefined optimization procedure, they are not prone to any subjective bias, instead featuring algorithmic objectivity. Furthermore, system states' weighting is implicit according to the cluster size, which allows for adequate representation of both extreme and rather common events in the approximation. Even though the implementation comes at higher costs compared to other approaches, they are superior when aggregating temporal resolution [7], and central in related research in recent years [1].

2.2. Research focus specification

The right-hand panel of Figure 1 illustrates a selection of relevant literature applying cluster analysis for temporal aggregation in electricity market modeling. It can be divided into three groups: hierarchical, partitional, and new techniques. They are briefly discussed in the following, with a detailed overview in Table 1.

Table 1: Characteristics of reviewed literature.

	Algorithm	Distance metric	cluster representation	aggregation level	# of clusters	cluster variables	application field
	HACAL						
[2]	WARD	Euclidean distance	closest to centroid	representative days	1,5,10	onshore wind, solar PV, electricity demand	Power plant investment and dispatch problem
[8]	WARD (based on [2])	Euclidean distance	centroid, medoid	representative days	5, 10, 15, 20	wind, solar, demand	Great Britain capacity expansion
[21]	WARD (based on [2])	Euclidean distance	closest to centroid	representative days	648 time slices in 54 days	wind, solar, demand	Power plant investment and dispatch problem
[23]	hierarchical (based on [2]), averaging	Euclidean distance	depending on algorithm	typical days	increasing number	GHI, temperature, wind	CHP based residential energy system, heatpump and PV based residential energy system, PV and wind based island system
[7]	Heuristics, WARD, Random Selection combined with optimal weighting, MILP Optimization	not specified	depending on algorithm	representative days	increasing number	electricity demand, onshore wind, solar PV	Generation expansion problem
[22]	WARD	Euclidean distance	closest to centroid	hours, days, weeks	increasing number	load, wind, solar profiles	Electricity capacity planning problem
[25]	WARD	Euclidean distance	closest to centroid	representative hours, days, weeks	4 weeks, 28 days, 672 consecutive time periods	renewables availability, demand	Capacity expansion planing with storage
	PARTAL						
[7]	Heuristics, WARD, Random Selection combined with optimal weighting, MILP Optimization	not specified	depending on algorithm	representative days	increasing number	electricity demand, onshore wind, solar PV	Generation expansion problem
[30]	k-Means	Euclidean distance	adjusted averaging	representative days	5 to 25	electricity demand	Power plant dispatch problem
[23]	k-Means, k-Medoids, hierarchical, averaging	Euclidean distance	depending on algorithm	typical days	increasing number up to about 70	GHI, temperature, wind	CHP based residential energy system, heatpump and PV based residential energy system, PV and wind based island system
[29]	k-Means	not specified	not specified	typical periods,	7 segmented time periods with in total 34 time steps	heating profile, solar irradiation, electricity demand, ambient temperature	Multi period district energy system optimization
[28]	k-Means	Euclidean distance	closest to centroid	operational states	increasing number	moments of overloads and net-loads	Transmission expansion problem
[27]	k-Means	Euclidean distance	medoid	snapshots	increasing number	nodal price differences, uncontrollable demand and generation, statistical values	Transmission expansion problem
[8]	k-Means	Euclidean distance	centroid, medoid	representative days	5, 10, 15, 20	wind, solar, demand	Great Britain capacity expansion
[34]	k-Means	Euclidean distance	centroid	representative days	different configurations	electrical, heating, cooling, gas load	Capacity configuration optimisation for regional energy system
[31]	k-mean, k-median, k-Medoids	Euclidean distance	depending on algorithm	typical daily profiles	increasing number up to 25	thermal and electricity demand, solar irradiation	Building energy system model
[32]	k-Medoid	not specified	not specified	typical consecutive periods	increasing number	building upon Kotzur et al. [23] (GHI, temperature, wind)	Seasonal storage modelling

	Algorithm	Distance metric	cluster representation	aggregation level	# of clusters	cluster variables	application field
[33]	k-Medoid	Euclidean distance	medoid	representative days	nine days	wind, solar, demand	capacity expansion problem with exogenous renewable investment
[35]	k-Means, k-Medoids, k-MILP	Euclidean distance	depending on algorithm	representative days plus extreme days	six days each	heat, cooling and electricity demand, prices, temperature, irradiance	capacity expansion and dispatch problem for multi-energy system
	non-traditional						
[36]	switching algorithm, k-Means, GMMs, NPMMs, different MS models	not specified	not specified	typical regimes	increasing number	simulated wind data	Identifying wind regimes
[37]	k-Medoids	band distance	medoid	typical days	fixed number	wind data	studying wind speed behaviour
[24]	k-Means, k-Medoids, DBA, k-shape, WARD	depending on algorithm	depending on algorithm	typical days	increasing number up to 9	electricity price	battery storage optimization, gas turbine dispatch
[26]	systematic sampling, k-Means, k-Medoids, WARD, Moment matching	depending on algorithm	depending on algorithm	representative days	not specified	load, on- and offshore wind,solar, and hydro availability data	Long-term transmission expansion model

Nahmmacher et al. [2] device a novel approach based on the hierarchical WARD algorithm to identify representative time-slices aggregating both demand and VRE time series. One time-slice contains three consecutive hourly system states. Their approach selects 48 time-slices, equivalent to six typical days, which are applied to a long-term capacity planning model of the European electricity system. Even though the advantages of HACAL over PARTAL are briefly discussed, no thorough reasoning for the selection of WARD among all HACAL is provided. Further applications of WARD ([8, 21, 23]) draw upon the approach by Nahmmacher et al. [2], where they use 648 time-slices and six to eight typical days. Yet, they do not motivate the choice of WARD. Poncelet et al. [7] evaluate multiple methods varying in complexity for identifying representative system states of demand and VRE time series, e.g., based on heuristics, WARD, or a mixed-integer linear programming optimization. They test the aggregation quality for the Belgian electricity system with 24 two-hour time steps. Again, the authors do not reason the selection of WARD. Merrick [22] employs WARD investigating the representation of temporal variability in electricity capacity planning models. The author states that WARD features clusters with observations that are extremely similar to each other or duplicates. Yet, the study lacks a reasoning for WARD in the context of other available algorithms of the same or other cluster classes. Pineda shows [25] that WARD allows the consideration of consecutive time periods. Although thoroughly examining how inappropriate aggregation of temporal resolution can induce substantial distortions of model output, and how this can be avoided, the line of reasoning for the selection of WARD is limited to a cluster class specific argument: HACAL do not require a set of k initial starting points, as it is the case for PARTAL, which the author considers advantageous. The author finds that it takes up to 300 representative days to properly capture variability induced into the system by demand and VRE fluctuations. Pfenniger [8] compares WARD, k-Means, and further heuristics to reduce input data for a capacity expansion for Great Britain. The author does not explain why these cluster algorithms from the set of available algorithms of the class of HACAL and PARTAL are selected.

The second subcategory contains temporal aggregation applications based on PARTAL. Green et al. [30] use k-Means to identify six to ten representative days for Great Britain's electricity demand net of wind generation, each containing 24 hourly system states. The authors argue that k-Means, besides being widely used and simply to implement, tends to form equally sized clusters, thus impeding the formation of system states that represent rare extreme system states. However, long-term investment models under- or overestimate the role of VRE if their variability including extreme cases is not properly accounted for [2]. Thus, their argument for the selection of k-Means over other cluster algorithms is disputable. Kotzur et al. [23] compare, *inter alia*, the performance of k-Means and k-Medoids for generating typical days and weeks. They stress that despite their greedy nature, these algorithms are relatively fast compared to hierarchical algorithms. Fazlollahi et al. [29] employ k-Means in combination with a parametric optimization to identify a minimum number of 34 representative periods without quality losses. They base their choice of k-Means solely on the frequency of its application in the literature. Similarly, Fitiwi et al. [28] lack a well-founded selection of k-Means in their implementation of a new method for the creation of 15 to 40 representative system states that adequately consider uncertainties in network expansion. Agapoff

et al. [27] conduct extensive sensitivities to investigate how the consideration of different features in the clustering analysis configuration affects model outputs of a classical transmission expansion problem. They find an optimal solution with at least 15 clusters. However, the motivation for the choice of k-Means remains unclear. Further research applying partitional cluster algorithms has been dedicated to related topics, such as temporal aggregation of data on building energy systems to find typical daily profiles [31], or the requirement of consecutive typical time periods for appropriate storage representation in energy system modeling [32]. Scott et al. [33] find that with reduced input data storage is used for addressing ramping challenges rather than temporal smoothing of VRE generation. They apply k-Medoids, arguing that it is less prone to smoothing than k-Means, thus giving better representations in long-term planning models. No further reasoning for the use of PARTAL over HACAL is given.

The third subcategory refers research applying recent, non-traditional cluster techniques. Kazor and Hering [36] compare 10 model-based cluster algorithms to find typical wind patterns from simulated wind data. Also clustering wind data, Tupper et al. [37] introduce a novel distance measure, the band distance. In contrast to the traditional Euclidean distance, this distance metric accounts for the shape of the time series, i.e., the relative distances of observations to other members of the data set. Teichgräber et al. [24] compare WARD and four PARTAL for a storage unit and a gas turbine dispatch optimization problem. The results suggest that the performance of the clustering method strongly depends on the purpose and the normalization of the data and the target function. A clear preference for an algorithm can therefore not be determined prior to the analysis. Härtel [26] comes to comparable statements. In addition to WARD and two PARTAL, the author also considers systematic sampling and moment-matching, which are combined with different scaling options for the input data.

Conventionally, two cluster classes, HACAL and PARTAL, seem to prevail in electricity market modeling literature. Per class primarily only one particular algorithm is applied in the aforementioned research: WARD or k-Means. However, the reviewed literature in the first and second subcategory lack thorough justification as to why the respective cluster class has been chosen, and why the applied algorithm has been selected over others within the same class. There is no systematic and sound evaluation of a large number of different hierarchical or partitional cluster algorithms in the context of temporal aggregation in electricity system modeling. Consequently, unintentional and undetected distortions of model outcomes may result due to the selection of a concrete cluster algorithm. The aim of this paper is to close this important gap in literature and scrutinize the common application of WARD and k-Means. Section SI.1 provides a brief introduction to cluster analysis focusing on HACAL and PARTAL. Research grouped into the category 'new techniques' offers potential for technical improvement of clustering applications in electricity market modeling, yet is beyond the scope of this paper.

3. Research design and fundamental input data

This section details the three-stage fundamental data clustering and market modeling approach. It further introduces the scenario variations of the conducted cluster analyses and defines metrics for as-

sessing the approximation quality of identified clusters.

3.1. Three-stage fundamental data clustering and market modeling approach

Figure 2 illustrates the three-staged approach for identifying representative system states. First, relevant input data are clustered using a large variety of cluster analyses, producing a large number of preliminary representative market scenarios. To compare results, stage two and three apply error metrics, which aim at identifying methods that generate the most representative, yet minimal, set of hourly system states.

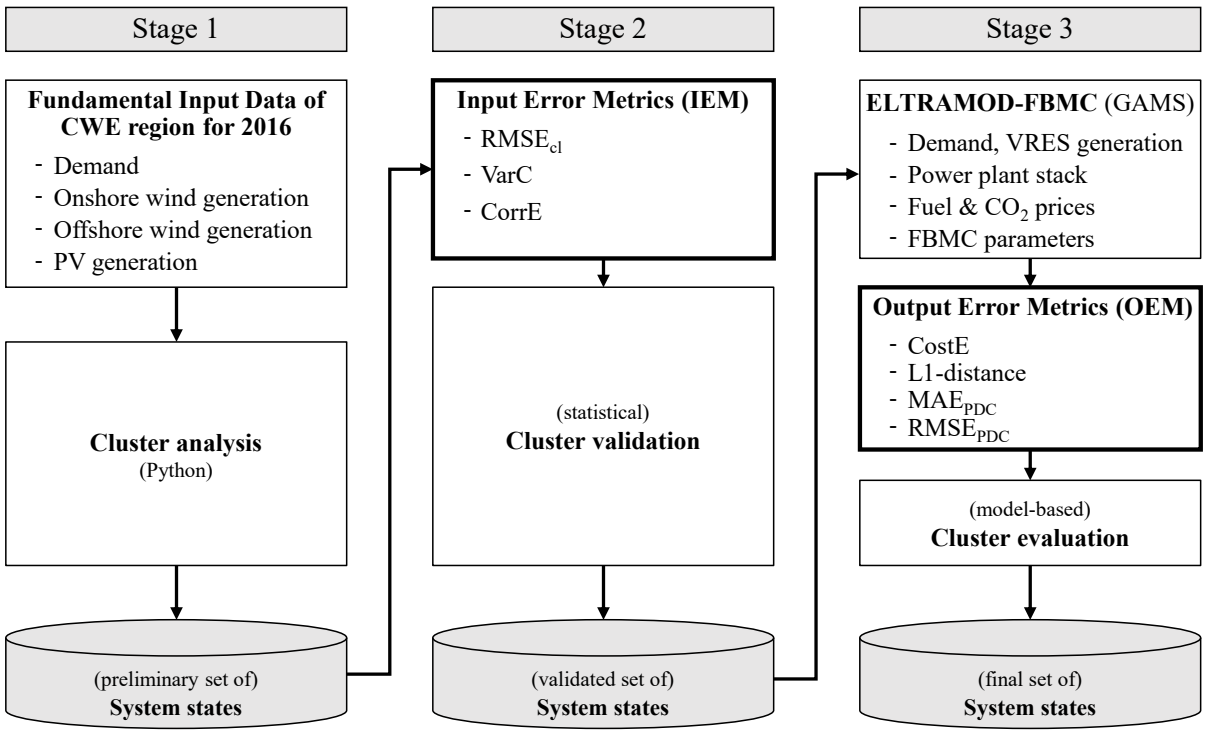


Figure 2: Three-staged approach for identifying representative system states.

3.2. Cluster analysis

System variables that induce variability into the electricity supply system, i.e., VRE generation and demand patterns, characterize a representative system state. Dispatchable generation technologies can adjust to the system variability. Hence, they are irrelevant for identifying representative system states. In this paper, electricity demand (hereinafter called load), onshore and offshore wind power, and PV for the CWE countries Austria, Belgium, France, **DE!** (**DE!**), Luxembourg, and **NL!** (**NL!**) from 2016 are clustered. This yields a multi-dimensional pattern matrix representing multiple system variables for multiple model regions with the dimensions $t \times f$, where $f = 19$ time series comprising $t = 8784$ uniformly sampled, continuous hourly values. The original data are retrieved from the Transparency Platform of the European Network of Transmission System Operator for Electricity. Data refinement is necessary due to incomplete time series (see Section SI.2 for a brief description and visualization of the refinement procedure). The data are re-scaled to the range $[0, 1]$ using the min-max normalization.

The investigated cluster scenarios differ regarding three methodological dimensions, as depicted in Figure 3. The first dimension relates to a total of ten different cluster *algorithms* (seven HACAL, three PARTAL). The second cluster dimension *representation* features a comparison of approximations based on both centroid-based and medoid-based representation. A centroid is the arithmetic mean of all observations in one cluster. A medoid is the observation that is closest to the centroid (see Section SI.1 for detailed elaboration). The third dimension is dedicated to the determination of the optimal *number of clusters* k, with a pre-defined range of 13 different $k \in Range_k = (1, 25, \ldots, 300)$.[3] In total, the three-dimensional cluster space yields 260 scenarios (10 algorithms $\times$ 2 representations $\times$ 13 different $k \in Range_k$).

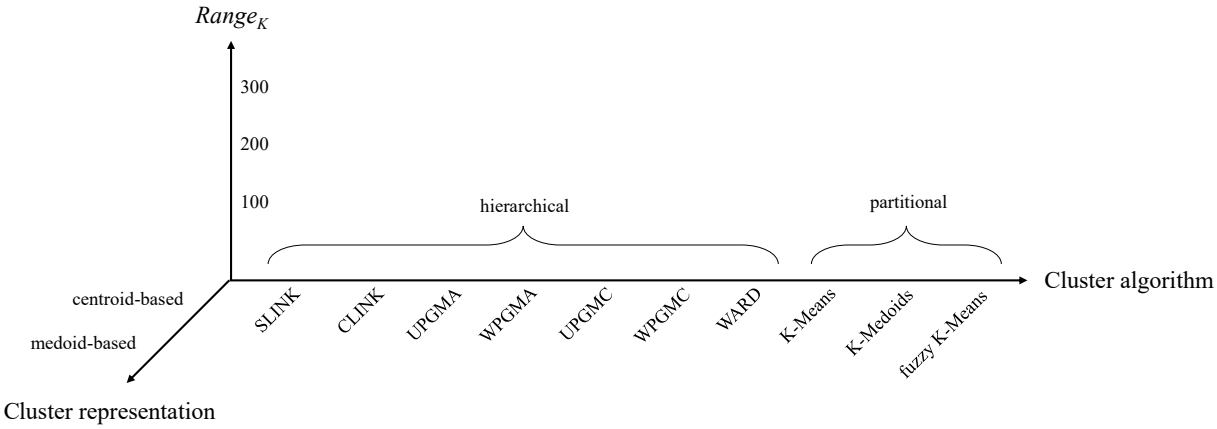


Figure 3: The three scenario dimensions algorithm, representation and chosen number of clusters.

3.3. Input data-based evaluation of cluster methods

The quality of cluster outputs can be assessed with either up- or downstream information of the clustering process [38]. The idea of cluster validation refers to an 'idealistic' analysis of the structure of the approximation compared to the structure of reference data set. More specifically, it is tested for accuracy and to what extent the obtained solution reflects the global statistical characteristics of the reference data set, e.g. temporal and spatial variability. Corresponding methods are referred to as Input Error Metrics (IEM). The term *input* relates to the upstream focus of cluster validation, as its scope is limited to the deviation between the clustered and the original input data set.

Each of the generated 260 cluster scenarios are validated with three *absolute* IEM. These metrics ensure that relevant temporal aspects of the original time-series are appropriately represented [2, 7, 21, 23]. Section SI.3 provides a detailed definition of all IEM. The Average Root Mean Square Error for Clustered Data ($RMSE_{cl}$) reflects the deviation from feature values of the original data set to their corresponding approximated representatives [2]. The smaller this metric, the closer resemble cluster representatives original values on average, and the more accurate is the approximation. The Covered Variance (VarC) reflects the grade to which the approximation captures the temporal variability of an original feature

[3] An incremental size of 25 is chosen to limit the amount of scenarios to be investigated to a feasible extent.

time series [17]. It is limited to the domain $VarC \in [0, 1]$. The closer it reaches the value of 1, the higher the average share of covered temporal variability in the approximation. The Correlation Error (CorrE) tests to what extent the correlation between the original time series is inherited to the approximation [2, 7], with CorrE $\in [0, 2]$: A CorrE of 0 indicates perfect representation, whereas a value of 2 reflects no representation of correlation at all.

To allow for meaningful representation, the three-dimensional scenario space is broken down to a new view based upon only two dimensions. Algorithm-Representation Scenario (ARS) aggregate the dimension *number of clusters*, while still differentiating the remaining two dimensions *algorithm* and *representation* subject to a specific $k \in Range_k$. They average the sum of each metric obtained for all members of the $Range_k$. In total, the collection of all possible combinations of these two dimensions yields 20 super-ordinate scenarios (10 algorithms $\times$ 2 representation forms). As an example, WARD and the two cluster representation forms, centroid and medoid, can be combined to the two ARS, WARD-c and WARD-m. Each of which could again be decomposed into the number of clusters k resulting in the originally investigated scenarios.

Altogether, the 260 scenarios amount to 39,020 hourly system states (1,951 system states $\times$ 20 ARS) equivalent to 741,380 feature-specific single observations. The sheer number of scenarios and system states renders a thorough consideration of each and every scenario impossible. The following *relative* IEM, based on [7], overcome this issue: the Relative Root Mean Square Error for Clustered Data (RMSE_{cl}^{rel}), the Relative Covered Variance (VarC^{rel}), and the Relative Correlation Error (CorrE^{rel}). They reflect the average absolute metric overall 13 different cluster number configurations $k \in Range_k$. This allows for a comparison of the aggregation quality of each ARS, irrespective of the chosen number of cluster k. The interpretation and domain are equivalent to the absolute IEM.

3.4. Market model-based evaluation of cluster methods

Previously conducted research suggests a model-based investigation to evaluate cluster quality with respect to the downstream employment of the approximation [2, 7, 21–23, 30]. The idea is to benchmark the cluster solutions with respect to their application-based performance [38, 39]. Here, clustered system states serve as scenario input in form of reduced time series to an electricity dispatch model. Output Error Metrics (OEM) then benchmark, *ceteris paribus*, model *outputs* against a base case, which draws upon the reference data set containing all 8784 hourly data points. Section SI.3 provides a detailed definition of all OEM.

The System Cost Error (CostE) is a convergence indicator and reflects the ability of approximated system states to replicate the base case's Total System Cost (TC). The order of the representative system states used as model input does not necessarily reflect the chronology of events in the original data set. This may impact inter-temporal constraints on ramping or storage. The CostE evaluates the approximation regardless this likely disruption of order. Even though the CostE allows for a comparison of the objective value among scenario runs, individual cost components and cost drivers might differ significantly, while TC remain the same. The notion of the L1-distance overcomes this issue. It decomposes the

objective function into its elements and subsequently sums the absolute deviations of the isolated terms of scenario solution values from the base case. It thus facilitates a comparison of individual cost drivers, such as curtailment of VRE [22].

Cluster quality can be further assessed by means of price formation comparisons [40]. The third and the fourth output error metric test to what extent clustered system states are capable of reproducing price formations. While the Mean Absolute Error of Price Duration Curves (MAE_{PDC}) represents the average difference between the Price Duration Curve (PDC) of a scenario and the base case, it underestimates strong deviations. Therefore, the Root Mean Square Error of Price Duration Curves ($RMSE_{PDC}$) is introduced, which is more sensitive to outliers. The same interpretation applies to all four indicators: The smaller the indicator, the closer are model results based on clustered time series to the base case. Both *absolute* and *relative* OEM are employed.

3.5. Electricity market model definition

Both the cluster analysis and aforementioned data refinement procedures are implemented in the open source programming language, Python. The employed Electricity Transshipment Model - Flow-Based Market Coupling (ELTRAMOD-FBMC) is formulated as linear optimization problem. It minimizes total system costs and covers the CWE region and adjacent countries. ELTRAMOD-FBMC draws upon the base model ELTRAMOD-INVEST [40] and has been enhanced to incorporate the incumbent cross-border transmission capacity allocation mechanisms - Flow-based Market Coupling (FBMC) and the Available Transfer Capacity (ATC) regime - in place in the considered electricity markets.[4] The ex post scope of this paper is limited to the CWE electricity system of 2016. The model is calibrated to replicate the historical electricity market outcome in 2016 in the CWE region in the base case. Capacity additions throughout 2016 are carried out exogenously. The cluster scenario-specific reduced multivariate time series represent exogenous demand and VRE generation information of CWE countries. All other fundamental input data, such as fuel prices and hourly heat demand, remain unchanged. Due to the non-consecutive character of the hourly system states to be found, any diurnal structures in the underlying data are disrupted by the clustering procedure. Model variables relevant to inter-temporal constraints, such as ramping or storage, suffering from this restriction, have been changed to exogenous parameters that reflect endogenous model results from the base case. Specifics of the implementation of the cluster analyses and model modifications are provided in Section SI.5.

4. Results and discussion on cluster analysis

This section discusses results on the statistical approximation quality of the identified clusters and related implications for the analysis.

[4] A detailed description of an NTC-based version of ELTRAMOD is provided in Ladwig [41]. An extended implementation of FBMC into the model is detailed in Schönheit et al. [42].

4.1. Results on absolute and relative input error metrics

Figure 4 illustrates the obtained absolute IEM.[5] As the number of clusters k increases, the global characteristics of the reference data set are better represented by the approximation. This is because, on average, a cluster representative needs to typify a lower number of observations grouped together in one cluster. In total, an increasing number of cluster representatives captures more information and variability. In general, while the obtained RMSE_{cl} and CorrE diminish with increasing k, the share of captured variability, VarC, rises. Single-Linkage Clustering (SLINK)-based ARS are outliers and suffer from weak approximation performance in terms of RMSE_{cl} and VarC. Intriguingly, absolute IEM incur diminishing returns to increasing k. Up to a cluster size of roughly 100, there is a significant marginal improvement of additional clusters to the approximation quality. Above this threshold, the marginal gain of additional clusters converges to near zero.

Table 2 aggregates absolute IEM over the original $Range_k$ and presents obtained relative IEM. WARD-c and k-Means-c achieve lowest RMSE_{cl}^{rel} scores among the investigated hierarchical and partitional ARS, respectively. Centroid-based ARS dominate the RMSE_{cl}^{rel} ranking, suggesting that centroid-based approximations closer resemble the reference data set (see also outperforming solid lines in Figure 4a). In contrast, medoid-based approximations show best aggregation performance in terms of average covered temporal variability and correlation error (see also outperforming patterns of the dashed lines in Figure 4b and 4c). WARD-m and k-Means-m are most effective here.

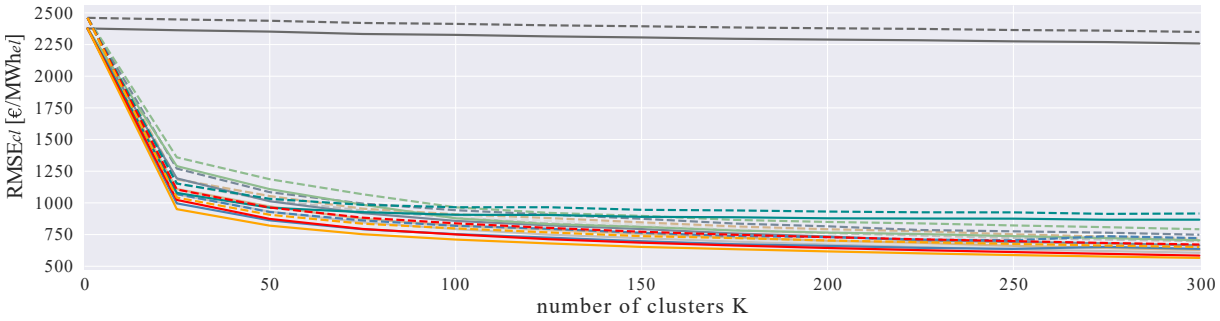
4.2. Discussion and implications for cluster dimensions

For the dimension *number of clusters*, results show that the approximation quality correlates positively with the number of clusters k. This finding is intuitive, yet reveals a fundamental trade-off and adverse relation of the two overarching requirements of the research objective: Finding a *representative*, yet *minimal*, set of system states. The desired minimal number of system states comes at the cost of accuracy of the approximation. Additionally, the analysis reveals that IEM incur the law of diminishing returns by increasing k. There are considerable marginal gains up to the threshold $k = 100$, whereas they decline after. To comply with the requirement *minimal*, while meeting the requirement *representative*, more prominence is allocated to the $Range_k = (1, \ldots, 100)$, whereas less concern is given to the cluster scenarios with $k > 100$ in the following. Results presented in Section 5 are based on the adjusted $Range_k^{'}$, which has an incremental size of five in $Range_k^{'} = (1, 5, \ldots, 100)$, and 50 in $Range_k^{'} = (100, \ldots, 300)$, totalling 25 different $k \in Range_k^{'}$.[6]

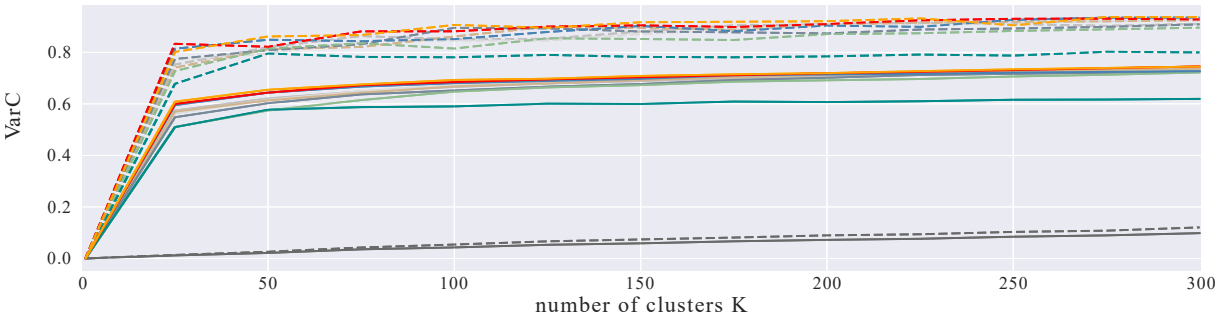
Regarding the two cluster dimensions *algorithm* and *representation*, four conclusions can be drawn. First, WARD is the most potent algorithm among the HACAL. A possible explanation may be that the

[5] Note that, each curve is not a continuous function but maps discrete values that are linearly extended in intermediate values.

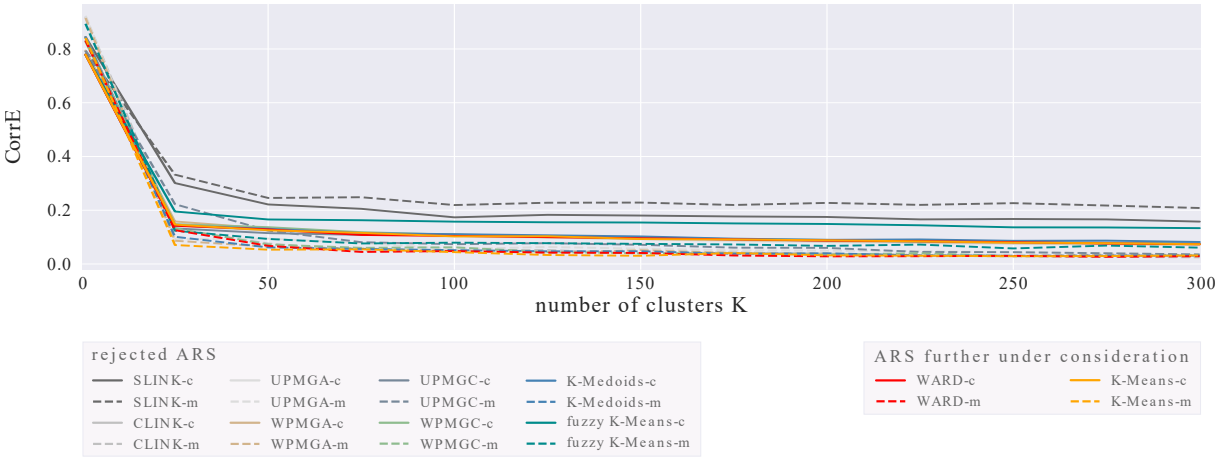
[6] Noteworthy, previous studies use the elbow method to determine the optimal number of k [5, 28, 30]. For the adjusted $Range_k^{'}$, no distinct curve kink can be determined, nor are the curves monotone. Thus, the elbow method is not applicable here.



(a) RMSE$_{cl}$ for all ARS over k.



(b) VarC for all ARS over k.



(c) CorrE for all ARS over k.

Figure 4: Overview of IEM for all ARS depicted over the original $Range_k$.

Table 2: Relative IEM RMSE_{cl}^{rel}, VarC^{rel} and CorrE^{rel} for 14 hierarchical and six partitional ARSs.

HACAL	RMSE_{cl}^{rel}	HACAL	VarC^{rel}	HACAL	CorrE^{rel}
WARD-c	842	WARD-m	0.8264	WARD-m	0.1059
CLINK-c	874	WPGMA-m	0.8082	WPGMC-m	0.1124
UPGMA-c	898	CLINK-m	0.8055	WPGMA-m	0.1135
WPGMA-c	904	UPGMA-m	0.8034	CLINK-m	0.1160
WARD-m	928	UPGMC-m	0.8018	UPGMA-m	0.1254
UPGMC-c	939	WPGMC-m	0.7815	UPGMC-m	0.1330
CLINK-m	964	WARD-c	0.6435	UPGMA-c	0.1485
UPGMA-m	979	UPGMA-c	0.6338	WARD-c	0.1502
WPGMC-c	981	WPGMA-c	0.6304	WPGMA-c	0.1518
WPGMA-m	991	CLINK-c	0.6296	CLINK-c	0.1519
UPGMC-m	1020	UPGMC-c	0.6196	UPGMC-c	0.1546
WPGMC-m	1064	WPGMC-c	0.6076	WPGMC-c	0.1570
SLINK-c	2311	SLINK-m	0.0671	SLINK-c	0.2399
SLINK-m	2399	SLINK-c	0.0546	SLINK-m	0.2781
PARTAL	RMSE_{cl}^{rel}	PARTAL	VarC^{rel}	PARTAL	CorrE^{rel}
k-Means-c	810	k-Means-m	0.8307	k-Means-m	0.1016
k-Medoids-c	855	k-Medoids-m	0.8169	k-Medoids-m	0.1086
k-Means-m	896	fzy-k-Means-m	0.7196	fzy-k-Means-m	0.1399
k-Medoids-m	926	k-Means-c	0.6474	k-Means-c	0.1508
fzy-k-Means-c	1023	k-Medoids-c	0.6373	k-Medoids-c	0.1554
fzy-k-Means-m	1082	fzy-k-Means-c	0.5494	fzy-k-Means-c	0.2015

algorithm merges the two clusters that are associated with minimum increase in inner-cluster variance. As much information as possible on the absolute observation values, as well as the temporal and spatial variability among the observations is preserved at each merging step. On the other hand, SLINK scores lowest ratings. This might be related to its tendency to form cohesive, elongated cluster chains [43]. Observations that are rather unalike are grouped together in one cluster causing a great loss of information.

Second, among all PARTAL, k-Means achieves salient results. This could be due to the employment of the enhanced k-Means++, which performs a more sophisticated seeding procedure before the cluster process starts [44]. It suffers less from its greedy nature to opt for local instead of global optima.

Next, ARS using WARD and k-Means outperform all other algorithms within their cluster algorithm class. The respective top-rated ARS obtain similar scores, indicating that neither of the two algorithms are superior to the other. Based on the relative IEM no straightforward decision in favor of either of the classes can be taken.

Finally, with respect to cluster representation, there are ambiguous results. A centroid-based cluster representation generates system states that closer correspond to their historical equivalents in the original data. This is because a centroid constitutes the arithmetic mean of a cluster. Every observation within a cluster equally contributes to the final representative. On average, the centroid represents all observations grouped together the best and the deviation from historical values altogether is minimized. However, extreme events are averaged out, causing a smoothing effect of the inherited variability. The drawback of centroids is the merit of a medoid-based cluster representation. Medoids inherit a greater deal of temporal and spatial variability. A medoid is an actual observation of the original data set and is more likely to contain extreme values, retaining a greater deal of the original variability.

In conclusion, IEM rule out inefficient cluster scenarios and allow for limiting the cluster dimension space.[7] The remainder of the analysis focuses on four ARS that proved most performant in approximation quality: WARD-c, WARD-m, k-Means-c, and k-Means-m, while applying the adjusted $Range_k^{'}$. The total number of scenarios reduces to 100 (25 different $k\times$ 2 representations $\times$ 2 algorithms).

5. Results on market model-based investigation

This section discusses results on the model-based approximation quality of the identified clusters and related implications for the analysis.

5.1. Results on absolute and relative output error metrics

The base case serves as a benchmark in the following elaboration. Its TC^{base} are 40.30 bn €, consisting of 40.03 bn € generation cost, 0.25 bn € curtailment cost, and further negligible cost components.

Figure 5 illustrates a selection of obtained absolute OEM. The curves using the same algorithm overlap almost perfectly with negligible deviations, regardless of the applied cluster representation. For simplicity,

[7] We assume that algorithms with inferior IEM results cannot achieve superior OEM results.

WARD is used interchangeably for both WARD-c and WARD-m as well as k-Means for k-Means-c and k-Means-m as of now. At the lower end of $Range_k^{'}$ remarkable incremental gains of additional clusters are realized. For approximately $k > 15$, the metric curves begin to oscillate with more or less strong kinks, and flatten at $k \geq 100$.[8] The latter does not allow for meaningful inference, but is caused by the discrete nature of of $Range_k^{'}$ that enlarges its increment size from 5 to 50 for $k \geq 100$. WARD-based curves have a rather smooth progression, whereas k-Means curves oscillate more extremely. If not stated otherwise, k-Means curves have their lowest local optimum at $k = 15$ and global optimum at $k = 70$ within the set of defined cluster numbers. Here, optima are either maxima or minima, depending on the considered metric. The lowest local and global optimum (also within the defined set of cluster numbers) of the WARD curves occur at $k = 20$ and $k = 55$.

The convergence behavior of TC^{ARS} toward TC^{base} is illustrated in Figure 5a. None of the TC^{ARS} reach the level of the base case, i.e., CostE > 0 over all $k \in Range_k^{'}$. Although achieved at different k, the optimal points across all ARS closely resemble and deviate by less than 0.35 bn €. The L1-distance curves closely follow CostE's equivalents and exhibit the same extrema properties.

Figure 5b shows generation costs, which never achieve the base case level but fall below by some 2.5 bn € on average. Since generation costs take the majority of the TC, the curves approximate CostE's patterns, only in a mirrored fashion. k-Means' global maxima within the defined set of cluster numbers at $k = 70$ considerably outperform WARD's by 0.5 bn €, whereas the optima at $k = 20$ (WARD) and $k = 15$ (k-Means) almost level.

Figure 5c presents VRE curtailment costs, which are mainly caused by must-run restrictions and excess VRE generation. The curves exhibit stronger oscillations, yet flattening the higher k. While WARD curtailment deviates rather moderately from the base case at $k = 20$, k-Means deviates largely at $k = 15$. At the previously considered extreme points $k = \{55, 70\}$, the difference declines significantly. The temporal occurrence of curtailment in scenario runs corresponds with the base case.

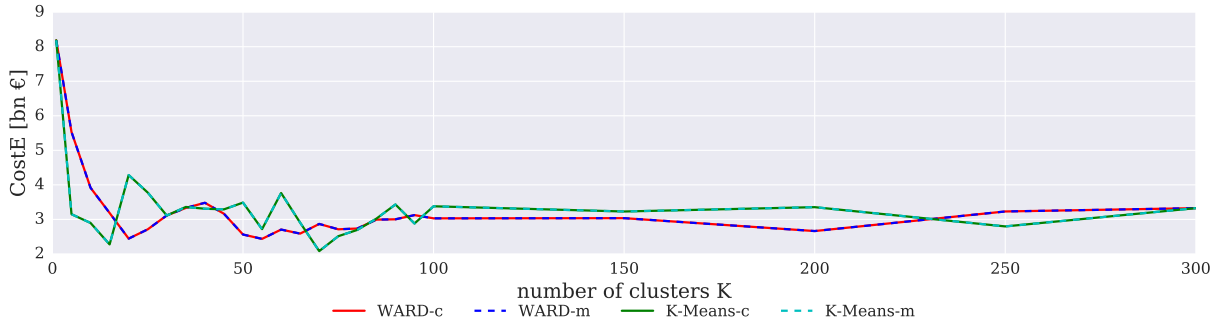
Price formation benchmarks are shown in Figure 5d. While $k = 20$ already constitutes WARD's global minima within the defined set of cluster numbers, k-Means' optima occurs at $k = 250$. Although obtained at very different k, minima levels are similar. WARD outperforms k-Means at lower $k \in Range_k^{'}$ and appears to run smoothly, whereas k-Means curves incur strong kinks. RMSE_{PDC} results are similar to MAE_{PDC} curves, and not illustrated here.

Table 3 presents obtained relative OEM. WARD-based ARS attain slightly better values across all relative OEM. As far as the representation forms are concerned, differences are insignificant.

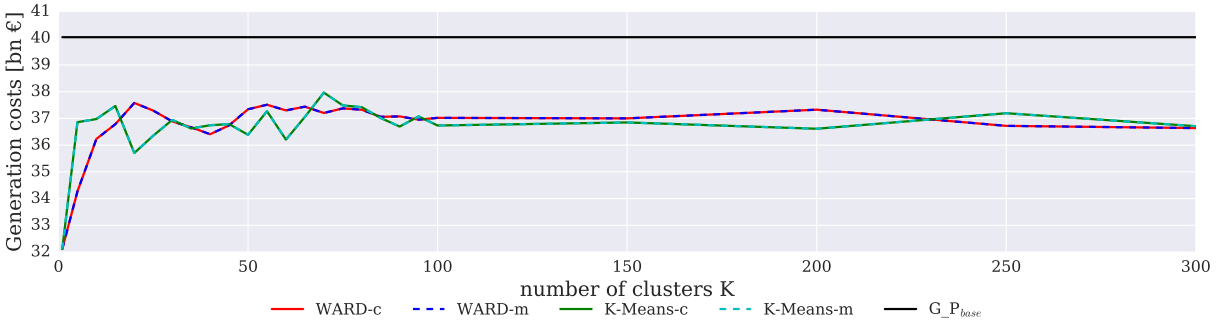
5.2. Discussion and implications for cluster dimensions

All OEM consistently support local or even global optima at $k = 15$ for k-Means and $k = 20$ for WARD (within the defined set of cluster numbers). Given the trade-off between the requirements *minimal* in

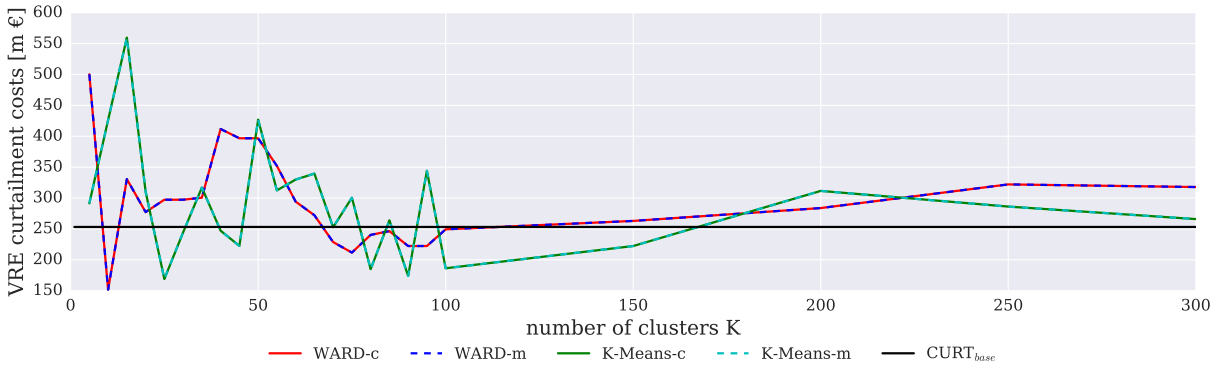
[8] Note that, for very large $k \gg 300$, the curves converge to zero in Figures 5a and 5d or to the base case indicated by the black lines in Figures 5b and 5c. Since this paper aims at finding a minimal set of representative system states, any $k \gg 300$ is not considered for the analysis.



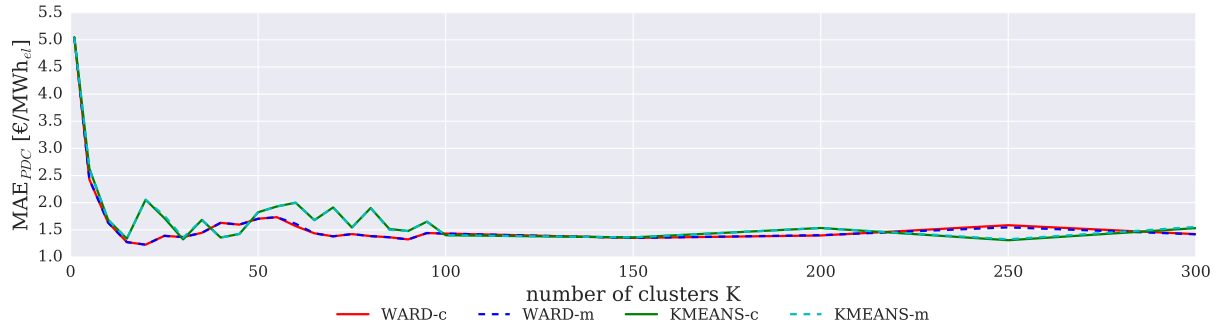
(a) CostE over k.



(b) Generation costs over k. The black line represents the base case level.



(c) VRE curtailment over k. The black line represents the base case level.



(d) MAE$_{PDC}$ for CWE countries in 2016 over k.

Figure 5: ARS view on the CostE, generation cost, VRE curtailment, and MAE$_{PDC}$ at solution. While not perfectly matching, the ARS curves using the same algorithm are very similar and overlap for many $k \in Range_k^{'}$.

Table 3: Relative OEM CostE^{rel} [bn €], L1-distance ^{rel} [bn €], MAE$^{rel}_{PDC}$ [€/MWh$_{el}$], and RMSE$^{rel}_{PDC}$ [€/MWh$_{el}$].

ARS	CostE^{rel}	L1-distance^{rel}	MAE$^{rel}_{PDC}$	RMSE$^{rel}_{PDC}$
WARD-c	3.283	6.731	1.6392	3.1584
WARD-m	3.283	6.731	1.6393	3.1557
k-Means-c	3.329	6.829	1.793	3.4170
k-Means-m	3.329	6.829	1.797	3.4239

number and *representativeness*, global optima at higher k are discarded. There are considerable gains in representativeness at lower ranges of $k \in Range_k^{'}$ up until the aforementioned lowest local optima since more variability in fundamental data can be processed by the model. Once the number of hourly system states is larger than the determined optimal thresholds, incremental returns on representativeness are not necessarily guaranteed. The fluctuations could be for two reasons. Perhaps the identified system states at locally (or globally) optimal k capture a great deal of temporal and spatial variability given their specific cluster parameters. Alternatively, the model happens to work well with the system states in conjunction with the time-dependent fundamental data of the respective time periods that have not been clustered. In conclusion, out of the 25 different $k \in Range_k$, $k = \{15, 20\}$ clusters are superior.

At optimal k, there is no cost convergence, neither for WARD nor k-Means. Instead, the investigated ARS similarly underestimate costs by some 2.5 bn € compared to the base case. One of the rare significant differences between WARD and k-Means appear in curtailment. In case of Denmark, as the major market zone suffering from curtailment, WARD outperforms k-Means at optimal k by far. In general, model shortcomings occurring in the base case also appear in scenario runs. Infeasibilities are directly caused by exogenous transmission and must-run restrictions that are not processed in the cluster analysis. However, clustered load and VRE generation patterns indirectly exert influence on these restrictions, which may force the model to curtail. The cluster analysis inherits temporal and spatial variability to the clustered system states, favorable or unfavorable in terms of model solution.

The analysis of price formations yields no significant difference between WARD and k-Means-based ARS. On average, the former appears to closer resemble the base case, yet only to a marginal extent. At optimal k, metrics results are similar.

Neither absolute nor relative OEM identify the most efficient cluster representation form. A centroid constitutes a synthetic mean of a cluster and has no affiliation to a certain point in time. This is why it is not possible to assign the remaining unclustered fundamental time-dependent input data to centroid-based system states. Instead, the hourly data of the observation that is closest to the centroid, i.e., the medoid, is used. This technical necessity is a small adjustment, yet may at least partially explain the almost perfect overlap of medoid- and centroid-based OEM.

To conclude, the model-based investigation allow for a further reduction of the cluster dimension space with regard to the optimal *cluster number*. While the WARD algorithm produces the most representative set of system states at $k = 20$, k-Means determines $k = 15$ as optimal. Due to the applied increment size

of five, it can be that the true underlying optimal $k \in [11, \ldots, 19]$ for k-Means and $k \in [16, \ldots, 24]$ for WARD. Given this uncertainty, the deviation of the optimal cluster number between the two algorithms is deemed irrelevant. As discussed in Section 2.2, the number of clusters determined in previous research ranges from 15 up to several hundred typical periods. The identified cluster number in this paper is located at the lower end of this range. Putting aside that WARD incurs less fluctuation of average cost deviation when varying k compared to k-Means, no recommendation in favor of either of the two algorithms and representation forms can be drawn.

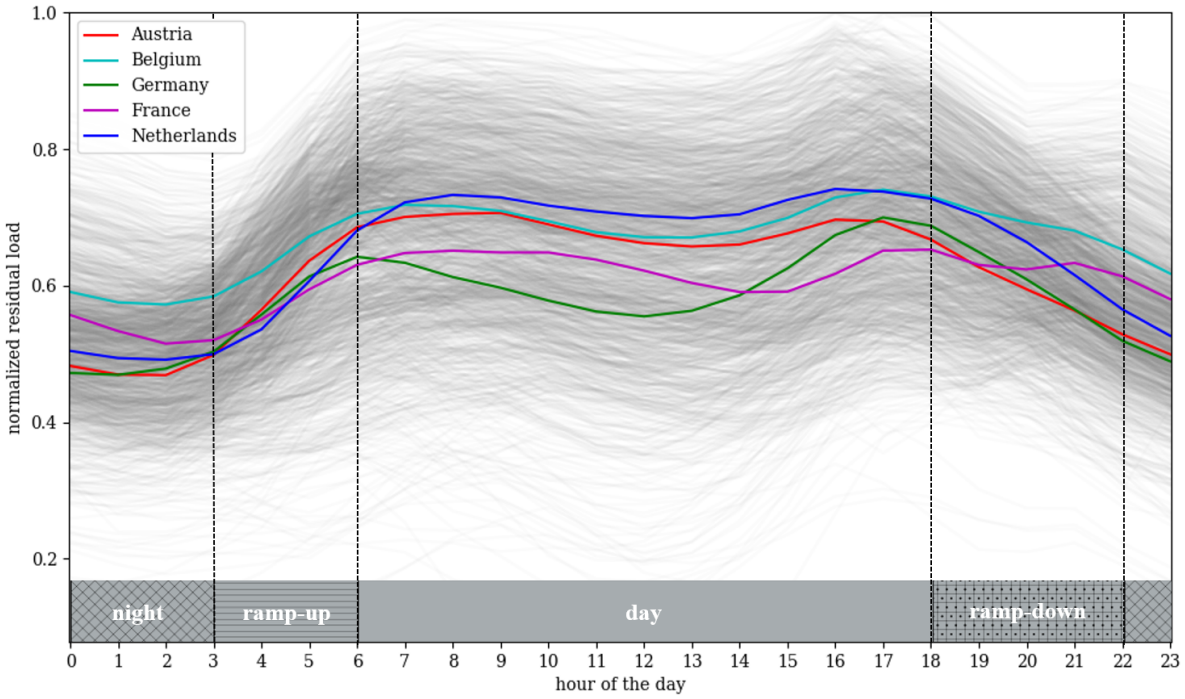
6. Analysis of the representative system states

Although load and PV follow typical diurnal patterns, wind penetration is rather random. To reconcile the patterns of all clustered system elements, Figure 6a shows a categorization of all hours of the day into four phases following diurnal fluctuations of residual load (load net of renewable generation) profiles of the CWE countries. The profiles resemble the typical 'camel back' with peak times during morning and afternoon.

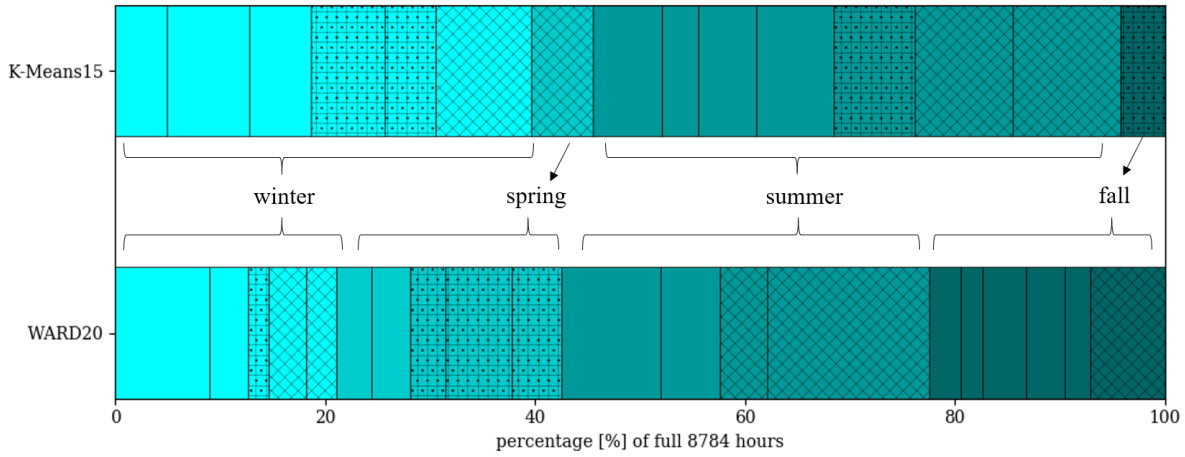
Figure 6b illustrates the seasonal and diurnal distribution of the identified representative system states. In terms of seasonal distribution, k-Means selects primarily system states from winter (six clusters) and summer time (seven clusters), representing approximately 40% and 50% of the original 8784 hourly observations, respectively. The spring cluster accounts for 6%, the fall cluster for 4%. In contrast, the system states chosen by WARD are rather evenly distributed throughout the year and across all seasons. Single cluster weights range from 2% to over 15% for WARD, and 3.4% to 10% for k-Means.

For diurnal patterns, both cluster configurations primarily select system states occurring during the day period: k-Means 47% and WARD 75%. The vast majority of these states are associated with hours of the morning hump and the subsequent saddle. Each cluster configuration selected only one hourly system state from the afternoon hump. The day period system states chosen by WARD are from all seasons. In contrast, k-Means picks only states from winter and summer days. Besides the day phase, WARD selects three spring evening ramp-down states, and five night states from summer, winter, and fall. k-Means identified four evening ramp-down and four night system states across all seasons. Intriguingly, none of the cluster configurations selected a hourly system state from the morning ramp-up.

Each system state is associated with a certain hour of the day, and contains a value for each each system element and country. Comparing each of these values to its average overall 366 values of this hour within 2016 allows for a coarse labeling of the system states with tendencies (Section SI.4 provides mathematical formulas for these averages). Table 4 illustrates these tendencies, which may be either largely positive ('high'), i.e., the levels of the representative system states tend to be higher than the average of the associated hours across one phase of the day. Or negative deviations ('low') prevail, i.e., the system state levels are below their average. A hyphen marks absent tendencies. For instance, WARD has the tendency to select system states with rather high load requirements during the day and night, and low load levels during the evening hours, with respect to the average load of the select hours of the phases.



(a) Daily residual demand profiles of the clustered time series and concomitant four phases of the day. Greyed lines depict all 366 daily profiles of 2016, colored lines show average profiles.



(b) Seasonal and diurnal distribution of the representative system states. Colors show seasonality, hatch patterns indicate diurnal structures: night period (10pm - 3am, cross-diagonal hatch), morning (3am - 6am, not visible as no morning hours are selected), daytime (6am - 6pm, no hatch pattern), evening (6pm - 10pm, dotted hatch with lines). The width of each cell corresponds with the cardinality of a cluster.

Figure 6: Diurnal phases of the residual load and seasonal and diurnal distribution of the determined system states.

These tendencies are generalizing, thus to be taken with caution since there are also few deviations contradicting the general tendency. The determined representative load values closely resemble their average for both k-Means and WARD. In contrast, deviations from their average range from 30 to 40% for wind, and centre around 80% for PV, without large differences between the cluster configurations.

Table 4: Tendency of the dynamics captured by the representative system states, compared to the hourly average feature values.

	WARD20			k-Means15		
	day	evening	night	day	evening	night
load	high	low	high	high	-	-
PV	-	high	-	high	high	-
wind onshore	-	low	low	low	high	low
wind offshore	-	low	low	-	-	low

The identified typical system states suffer from a smoothing of extreme events (see Section SI.1 for elaboration). Table 5 illustrates the Mean Absolute Error of the system states to the corresponding maxima (MAE_{MAX}) and minima (MAE_{MIN}) for each system element. They indicate the mean deviation of the values inherited to the system states from the maximum or minimum of the original data set of that hour a system state is associated with (see Section SI.4 for mathematical definition). By way of example, the hourly load levels of the identified system states are, on average, almost nine gigawatts below and above the corresponding maxima and minima of the hours associated with the selected states. This smoothing explains why there is no cost convergence towards the TC^{base} or the generation costs of the base case (Figures 5a and 5b). Lower upper and lower extreme load levels can be supplied with less cost-intense generation technologies. The representation quality of k-Means outperforms WARD with regard to capturing low load and wind generation extremes, as well as high load and high PV generation patterns. WARD incurs lesser smoothing for low PV and high wind generation patterns. However, differences between the two cluster configurations are rather marginal compared to the amplitude of the metrics.

Table 5: Mean Absolute Error to the corresponding maxima or minima of the hours associated with the system states for each system element $[MWh/h]$.

	k-Means15		WARD20	
	MAE_{MIN}	MAE_{MAX}	MAE_{MIN}	MAE_{MAX}
load	8790	8573	8857	8665
PV	1194	1293	1124	1802
wind onshore	2355	6523	2461	6321
wind offshore	675	1049	695	1027

7. Summary and modeling implications

This paper presents a structured three-staged approach to evaluate and compare hierarchical and partitional cluster methods for temporal aggregation in electricity market modeling. Fundamental electricity market data on electricity demand, PV, and wind penetration in the CWE region are clustered and serve as input to a flow-based electricity dispatch model. In total 260 cluster scenarios, differing in terms of cluster algorithm, cluster representation, and cluster size, are scrutinized by IEM and OEM. These metrics allow for statistical validation and application-based evaluation of cluster scenarios.

Findings support the selection of appropriate cluster algorithms for time series reduction in computational costly electricity market modeling. WARD and k-Means are most performant among hierarchical and partitional cluster algorithms. Even though k-Means patterns are slightly more sensitive to variations in the number of clusters, i.e., they incur more volatility, there is no significant difference between WARD and k-Means. This justifies the selection of WARD and k-Means in previous model-based analyses.

The devised approach is applicable to various types of electricity system model applications or, more generally, to applications where a temporal aggregation of time-variant fundamental input data is required. The selection of data that is being reduced by cluster analysis depends on the scope and technology portfolio of the used energy model. For instance, additional electric loads from other energy-intense sectors that are exogenously added to the demand from the power sector should be included. Likewise, demand profiles for driving from battery-electric vehicles should be clustered. In contrast, there is no need to cluster load profiles of endogenously modeled loads, such as endogenous charging and discharging of battery-electric vehicles with re-conversion option. Clustering weather data instead of VRE availability factors may increase approximation quality, yet, comes at the cost of increased complexity and computational burden as more variables need to be processed. Historical VRE generation profiles may underestimate the actual generation potential since they are already reduced by VRE curtailment. With an increasing electrification of the entire energy system, VRE surplus is likely to be largely used beyond the power sector in the future. Clustering residual load may resolve this issue, which merits future research.

This paper also provides an approach for assessing and comparing a large number of possible cluster analysis configurations. In light of the arising of many new aggregation techniques, the approach can guide future selection processes of appropriate temporal reduction methods.

Besides cluster algorithm, representation quality depends on selection of cluster representatives. Centroid-based approximations best resemble historical values, whereas medoids are more efficient in capturing both spatial and temporal variability of the original data set. Hence, medoid-based representatives are more suitable for applications that strongly rely on an appropriate representation of (co-)variance present in the original data.

The use of IEM reveals the law of diminishing returns with respect to the number of clusters. This means that the number of clusters, hence also the number of representative system states, can be substantially reduced. Here, a distinct threshold for an adequate number of clusters at $k = 100$ is determined.

A large part of the original 8,784 hourly observations can be covered with the significant smaller number of 100 representatives. OEM results even support a further reduction down to $k = \{15, 20\}$, depending on the algorithm applied. Although OEM results oscillate at smaller k, the average deviation between the unclustered benchmark and scenario outcomes generally remains at a similar, relatively low level. In other words, aggregation quality stays the same, regardless the precise number of clusters chosen for $k \in Range_k^{'}$ when used in short-term dispatch modeling. Hence, whenever an application's computational burden is high, clustering can be an efficient and effective method for coping with increasing model complexity within the discussed limitations, even for a very low number of representative system states.

The proposed methodology allows modelers and market participants to gain valuable insights into typically occurring system states, hence facilitating a better understanding of the underlying market functionality. Here, WARD selects representative system states from across all seasons mainly associated with day-time hours with high load levels, high PV and low wind penetration. In contrast, k-Means mainly picks winter and summer system states from day-, evening- and night-time, with high demand and PV levels and erratic wind penetration. This is relevant to applications that model dispatch and investment decisions for energy infrastructure with strong seasonality or diurnal structures in demand or supply, such as heat provision.

Certain applications of electricity system models strongly rely on extreme events. Time aggregation methods, by construction, exclude outliers from the approximation, thus exerting a smoothing effect on the clustered data. Long-term models capable of endogenous generation and transmission capacity additions, determine optimal investment decisions such that the maximum residual load can be covered. The absolute value of maximum residual load contained in the original data set and clustered data may differ, thus potentially impacting model outcomes. In this application, the degree of smoothing does not differ significantly between WARD and k-Means.

Noteworthy, the selection of OEM depends on the energy model that uses the clustered input data. For capacity expansion models, OEM should evaluate the approximation regarding the representativeness of endogenous capacity variables. Furthermore, results based on OEM must be assessed relative to the context of their application. Depending on the specific research question and model formulation, temporal aggregation of time series to hourly, non-consecutive system states may influence inter-temporal constraints, such as ramping and storage inter-temporality. Furthermore, time series from different years are likely to influence attained cluster solutions. This merits future research, which may concentrate on the generation of representative system states consisting of multiple, consecutive hours. Additionally, dynamic cluster algorithms may bear potential for improvement of approximation quality, in particular time-spanning flexibilities such as storage or demand side management.

Since we apply a dispatch model with a short-term focus, our analysis is limited to a data set comprising only one year. For long-term capacity addition models, using multiple weather years may improve robustness of cluster results and accuracy of IEM and OEM. Further, we use the Euclidean distance as proximity measure. Other proximity measures that put more weight on outliers may cause a variation in results. We leave these aspects to further research.

Energy 247 (2022)

Temporal aggregation of time series to identify typical hourly electricity system states: A systematic assessment of relevant cluster algorithms

Supplementary information

Martin Kittel, Hannes Hobbie, Constantin Dierstein

SI. Supplementary information

SI.1. Fundamentals of cluster analysis

Han et al. [45] describe data mining as a process to identify and extract useful knowledge in form of pattern recognition from a comprehensive data set, e.g. based on machine learning. The latter is dedicated to find ways for computers to learn automatically based on data and can be divided into two main disciplines: supervised and unsupervised machine learning [43, 46]. While supervised learning relates to the classification of unlabeled patterns based on already labeled training data, unsupervised learning corresponds to an exploratory learning process without already predefined classification semantics. Cluster analysis is a major discipline of unsupervised machine learning and is used for data segmentation. Attributes of sub-populations within unlabeled data can be identified in order to compress data. Instead of applying individual data points, obtained clusters can serve a specific purpose [43].

The generic procedure for pattern clustering follows five consecutive steps as illustrated in Figure SI.1 [39]: First, pattern representation is analyzed in terms of feature constitution of the unlabeled data, such as number, type and scale. Second, given quantitative features, proximity is defined as a distance metric that quantifies similarities of patterns. Third, a cluster algorithm is carried out. Identified structures in the data allow for the extraction of a simple and compact representation of the reference data. Finally, the obtained output requires a validity analysis to test for meaningful clusters. Once the cluster solution is validated, the user then interprets clusters.

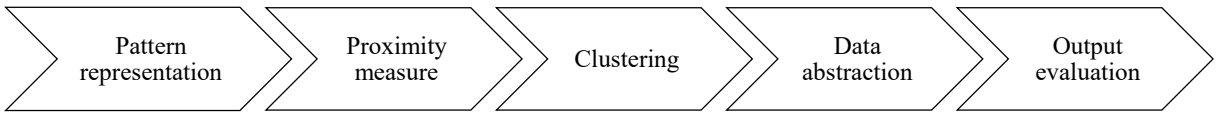


Figure SI.1: General clustering procedure.

Given quantitative features, proximity of two observations is indicated by a distance metric that quantifies (dis)similarities of patterns. Patterns can consist of single or multiple data points; hence, distance can be measured between single observations and/or groups of observations. In line with aforementioned studies, this work focuses on the most popular and easy-to-access distance metric, the Euclidean distance.

The research community has developed a rich variety of algorithms, on the order of thousands, since the 1960s [46]. *Inter alia*, they are distinguished by the separation of the

obtained output cluster structure, which can be hard or fuzzy [43, 46]. In fuzzy clustering, a pattern can belong to multiple clusters to some degree of membership, while hard clustering allocates each pattern to exactly one cluster. The partitioning criterion defines whether clusters are of the same semantic level or are hierarchically ordered, incorporating different semantics at each level of the hierarchy. Agglomerative HACAL perform 'bottom-up' merging of single observations until a termination condition is satisfied or all observations are grouped into one cluster altogether. Essentially, there are five categories for clustering analyses as depicted in Figure SI.2: Hierarchical, partitional, density-based, grid-based, and model-based algorithms [45–47].

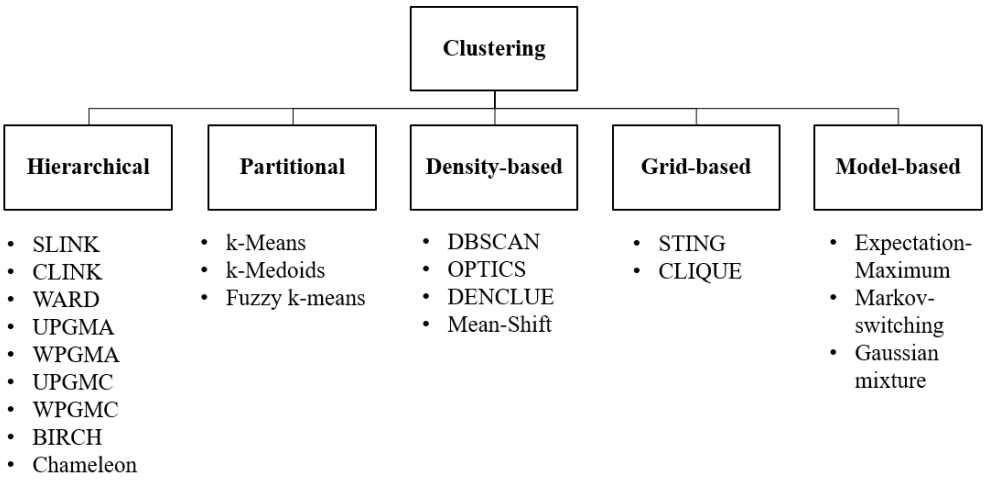


Figure SI.2: Taxonomy and categorized examples of clustering algorithms (based on [45–47]).

The generic procedure for agglomerative HACAL is carried out as follows [43]: Let N be a set of observations, D the distance matrix denoting the Euclidean distance d_{mo} between all observations or patterns $m, o \in N$, u_i the cardinality of a cluster C_i. As a first step $v = 0$, the algorithm places all observations into singleton clusters, i.e., one observation per cluster. The obtained universal cluster structure is denoted $\mathfrak{C}_v = \mathfrak{C}_0$. Based on the ordered list of distances for all distinct pairs of observations or patterns D_v, the algorithm finds the pair of clusters that is closest to each other, i.e., it minimizes distance D_v for all clusters C_k and C_j comprised in $\mathfrak{C}_v$. In case of a tie, it chooses any pair with smallest distance:

$$D_v = \min D(C_k, C_j) \; \forall k \neq j; C_k, C_j \in \mathfrak{C}_v \tag{SI.1}$$

This merging step is iterated until all observations are grouped into one cluster $N \in$

$\mathfrak{C}_{n-1}$. In order to keep computational costs at minimum, the distance matrix D can be calculated in a recursive way after each merging step [48, 49].

There are a vast variety of agglomerative HACAL. While they all follow the generic procedure in (SI.1), they differ in terms of the employed linkage measure that determines the clusters to be merged. It is denoted by the construction formula of the distance measure $D(C_k, C_j)$ and its recursive parameters. Seven agglomerative HACAL are employed in this paper: SLINK, Complete-Linkage Clustering (CLINK), Unweighted Pair Group Method using Arithmetic Mean (UPGMA), Weighted Pair-Group Method using Arithmetic Mean (WPGMA), Unweighted Pair-Group Method using Centroids (UPGMC), Weighted Pair-Group Method using Centroids (WPGMC) and WARD.[9]

The majority of these algorithms join clusters according to a linkage measure that aims at merging the most similar clusters. By contrast, WARD is based on the idea of minimizing the loss of homogeneity, i.e., the loss of information associated with each merging step [50]. This loss is quantified by the Sum of Squared Errors (SSE) objective function. To compute the SSE, the mean vector of all observations in C_k, called centroid $\bar{x}_k$, is required:

$$\bar{x}_k = \frac{1}{u_k} \sum_{n \in C_k} x_n \tag{SI.2}$$

The centroid $\bar{x}_{k+j}$ for a newly merged cluster C_{k+j} is defined by all original observations in C_k and C_j. The distance between the new cluster C_{k+j} and any other cluster is computed by the Euclidean distance between the centroids, respectively. WARD aims at minimizing its objective function SSE, which is based on the inner-cluster variance [50]:

$$SSE = \sum_{n \in C_k} ||x_n - \bar{x}_k||^2 \tag{SI.3}$$

The algorithm merges the two clusters that are associated with minimum increase in inner-cluster variance when merged. In other words, it compares all potential merging pairs, as well as corresponding deviations as measures for information loss and selects the merge that results in the smallest deviation from the new centroid.

In contrast to HACAL, which produce a cluster hierarchy, PARTAL obtain a one-level, non-hierarchical partition per clustering step. Their application requires three parameters

[9] Linkage measures and recursive parameters are detailed in [49].

of the cluster analysis procedure to be specified *a priori*: The number of desired clusters k, the cluster initialization procedure and the distance measure [46]. By way of example, the general partitional clustering procedure can be described based on k-Means - probably the most renown partitional cluster algorithm - which optimizes an objective function $J(C)$ based on SSE over all clusters determined in number by k [43, 46, 51]: At first, the algorithm places k initial centroid seeds $\bar{x}_k$ across the feature space or data, either randomly or according to some seeding criterion. It then assigns each observation to the cluster C_k with the shortest distance from the respective cluster centroid $\bar{x}_k$ to the observation, using some distance measure. In case of a tie, it chooses any cluster C_k with smallest distance:

$$J(C) = \min \sum_{K} \sum_{n \in C_k} ||x_n - \bar{x}_k||^2 \tag{SI.4}$$

Given the obtained partitioning is denoted $\mathfrak{C}_1$, the algorithms recalculates the centroid $\bar{x}_k$ as the mean of all members of cluster C_k. The reassignment is iterated such that new partitions $\mathfrak{C}_{v \geq 2}$ are generated until some convergence criterion is met. Typical criteria are no or only marginal relocation of cluster centroids; no or only marginal reassignment of cluster memberships; or some threshold for SSE decrease. Since it minimizes deviation between observations and respective centroids, the objective function $J(C)$ can be interpreted as cost of configuration. The procedure above implicitly ensures that, at each clustering iteration, the value $J(C)$ decreases.

HACAL lead to relatively low computational costs, as no combinatorial complexity needs to be considered [43]. However, they suffer from rigidity [45, 46]. Cluster structures found at each merging step cannot be undone and the next step might build upon an erroneous cluster partitioning. In contrast, PARTAL induce high computational costs as theoretically only an exhaustive enumeration of all possible combinations yields a global optimum. To overcome this complexity they adopt greedy heuristics. Instead of global optima, they opt for the locally optimal choice.

Density-based methods draw upon the number of observations in the dimensional neighborhood. Clusters are grown as long as the density exceeds some termination criterion, such as a minimum amount of observations within a certain radius. As for grid-based methods, a finite number of grid cells are produced covering the data space. For each cell some integrated clustering operation is conducted, which then can be combined with

methods of other cluster categories. Finally, model-based approaches assume that data were produced according to model structure and attempt to recover a model for each cluster. Once recovered, it then determines the assignment of the original data set to clusters [47].[10]

Once a final cluster partition is determined, a representative needs to be retrieved from any identified sub-population. It is derived by reducing the entire cluster to a single, typical profile that ideally accurately reflects the global characteristics of all cluster members [30]. A straightforward and intuitive approach is to use the centroid $\bar{x}_k$ as representative. Although simple averaging allows every observation within a cluster to contribute equally to the final representative, it causes the centroid to lack extremes, thus inducing temporal smoothing of extreme values. Instead of creating a synthetic representation, the actual observation that best represents the cluster can be used, i.e., the observation that is closest to the centroid, called the medoid $\hat{x}_k$ of a cluster C_k. It is more likely to contain extreme values.

$$\hat{x}_k = arg\min_{x_l} \frac{1}{N} \sum_{i=1}^{N} ||x_l - x_i||^2 \tag{SI.5}$$

Due to the fact that clusters differ in number of assigned observations, each representative needs to be weighted with respect to its relative cluster size.[11] This accounts for the fact that there are observations reflecting rather common system states, whereas extreme system states occur rather occasionally. Besides the weighting according to the relative cluster size, no further weighting is carried out.

SI.2. Data refinement procedure

The pattern matrix is not readily applicable for two reasons: First, the original data exhibit different temporal and spatial resolution. For instance, German VRE generation or demand time series were only retrievable with a 15 minute interval structure, while ELTRAMOD-FBMC requires an hourly resolution. The data are aggregated and distributed according to the resolution of the model.

Second, many of the time series are incomplete, requiring deliberate data refinement. Complete time series are an crucial prerequisite to cluster analyses. Missing data points

[10] As elaborated in Section 2.2, the aim of this paper is to scrutinize the usage of HACAL and PARTAL. Thus, density-, grid-, and model-based algorithms are beyond the scope of this paper.

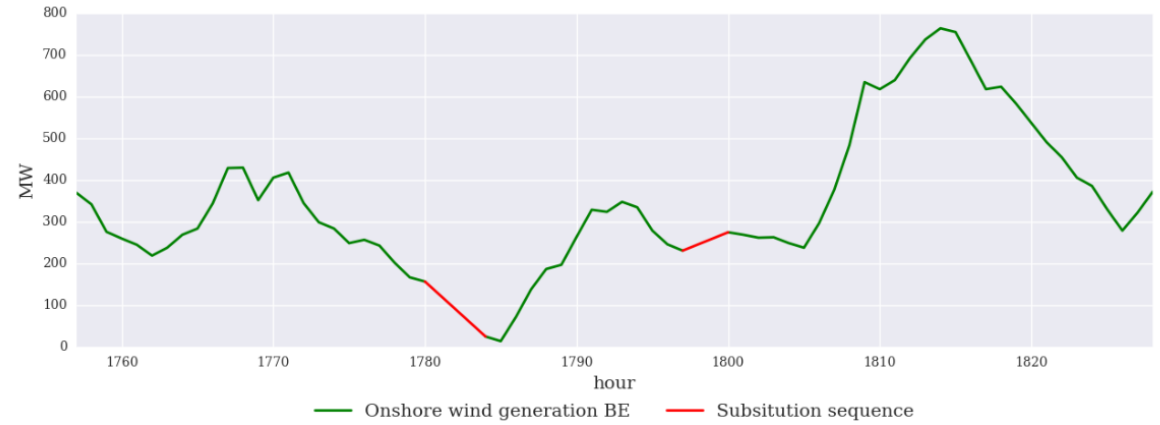
[11] See [2] for the applied weighting method in this work.

impede comparability of observations and computation of distances between one another, thus potentially distorting obtained cluster solutions. The data exhibit three categories of gaps: Up to three consecutive hourly data points, four to 24 consecutive data points, and gaps comprising more than an entire day. Each gap is replaced with a sequence of synthetically generated data points, which is tailor-made to the respective gap category and feature. Gaps of the first category are linearly interpolated, as illustrated in Figure SI.3a. Since load and VRE profiles underlie intra-weekly and, in the case of PV, diurnal patterns, the second category requires a more deliberate replacement procedure, which is displayed in Figure SI.3b. While each hour of the replacement sequence for load time series is derived from the mean of the same hour of the same weekday one week before and after, VRE replacement sequences are derived from the rolling average of the corresponding time section of the previous and the following day of the same time series. To avoid jumps in the data, each sequence is linearly factorized to fit the level of the time series at the respective gap, i.e., it is multiplied by the average level of the last and first existing data point that encompass the gap. While the same procedure applies to gaps in load data comprising more than 24 hours, the substitution of gaps within VRE generation time series follows a more sophisticated approach. Assuming that the large-scale weather conditions of neighboring countries closely resemble, VRE generation does so, too. This relation can be used to approximate periods of missing data, as shown in Figure SI.3c. The most suitable substitution area is assigned to each country and generation technology, according to geographical constitution or data quality. The replacement sequence is selected from the reference time series of the replacement area corresponding to the time section of the gap. To smoothly fit it into the time series containing the gap, the replacement sequence is scaled by the generation capacity ratio of the respective countries and technologies.

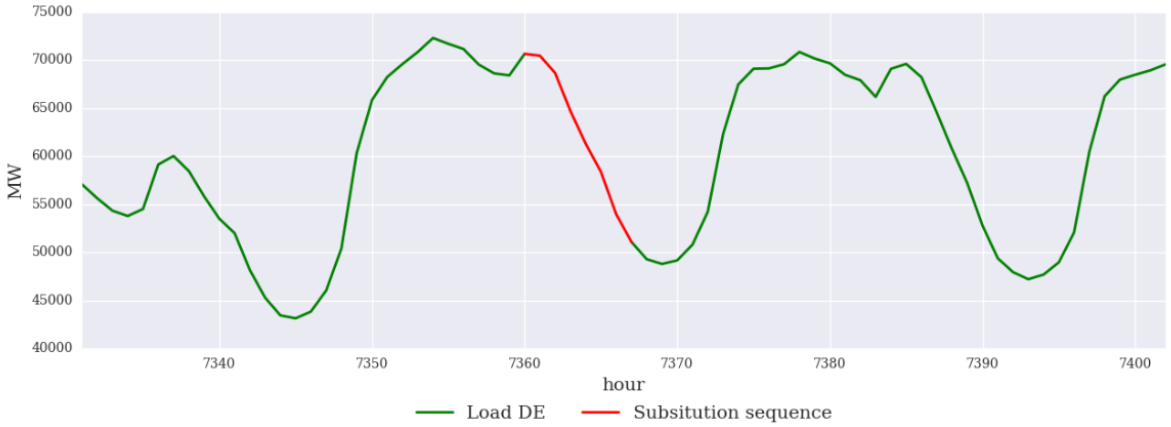
SI.3. Mathematical formulary for input and output error metrics

Let N be the total number of observations, f a feature, F the total number of features, C_k a cluster in the clustering partitioning $\mathfrak{C}$, $x_{f,n}$, $\bar{x}_{f,k}$ and $\hat{x}_{f,k}$ historical values and their corresponding cluster representatives, respectively. The RMSE_{cl} over all features and all hourly values is denoted as:

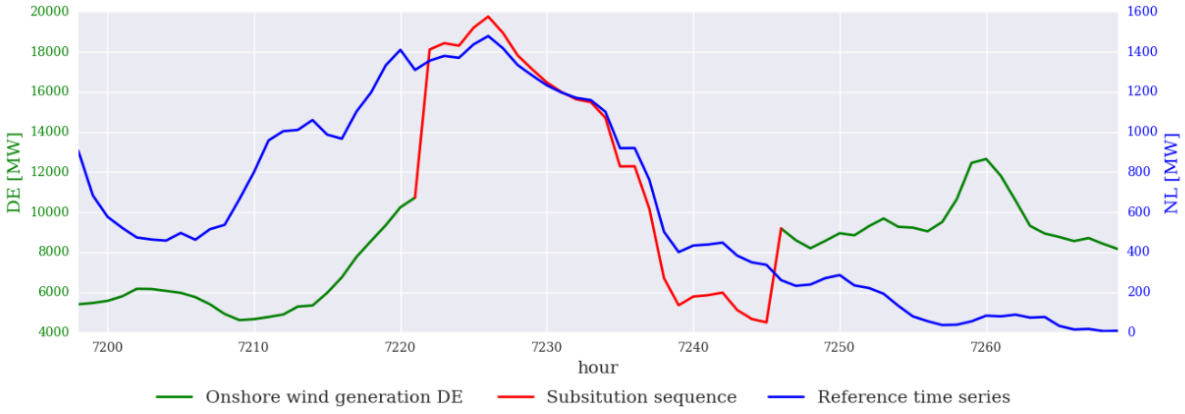
$$RMSE_{cl}(\bar{x}) = \frac{1}{F}\sum_{f\in F}\sqrt{\frac{1}{N}\sum_{C_k\in\mathfrak{C}}\sum_{n\in\mathfrak{C}}(x_{f,k}-\bar{x}_{f,k})^2} \tag{SI.6}$$



(a) Exemplary gaps in Belgian onshore wind generation data and respective substitution sequences. The gaps encompass three and two hourly values, respectively.



(b) Exemplary gap in German load data and substitution sequence. The gap encompasses six hourly values.



(c) Exemplary gap in DE onshore wind generation data and substitution sequence. The gap encompasses 24 hourly values.

Figure SI.3: Exemplary gaps in data for all three categories and respective substitution sequences..

Let f_s and f_s^* be a feature of the original and approximated data set based on either the centroid or the medoid, respectively. The VarC of over all feature time series is defined as:

$$VarC(f_s) = \frac{1}{F} \sum_{f \in F} \frac{var(f_s^*)}{var(f_s)} \tag{SI.7}$$

Let f_s and f_r be two features of the feature set and f_s^* and f_r^* their corresponding approximations. The CorrE is denoted as:

$$CorrE = \frac{1}{F} \sum_{f \in F} |corr(f_s, f_r) - corr(f_s^*, f_r^*)| \tag{SI.8}$$

Let TC_{base} and TC_{scen} be the TC of the base case and the model scenario, respectively, the CostE is defined as:

$$CostE = TC_{base} - TC_{scen} \tag{SI.9}$$

Let Z be the total number of market zones z, T the set of hours t, $p_{t,z}^{base}$ and $p_{t,z}^{scen}$ the day-ahead price in zone z at the hour t of the base case and of a scenario, respectively. The MAE_{PDC} and RMSE_{PDC} are then denoted as:

$$MAE_{PDC} = \frac{1}{Z * T} \sum_{z \in Z} \sum_{t \in T} |p_{t,z}^{base} - p_{t,z}^{scen}| \qquad \forall\ z \in Z_{CWE} \tag{SI.10}$$

$$RMSE_{PDC} = \frac{1}{Z} \sum_{z \in Z} \sqrt{\frac{1}{T} \sum_{t \in T} (p_{t,z}^{base} - p_{t,z}^{scen})^2} \qquad \forall\ z \in Z_{CWE} \tag{SI.11}$$

Let $card(\cdot)$ be the cardinality of a range, i.e., the number of elements contained in the range, the RMSE_{cl}^{rel} is then denoted as follows, while the remaining relative IEM and OEM are computed accordingly:

$$RMSE_{cl}^{rel} = \frac{\sum_K RMSE_{cl}}{card(Range_k)} \tag{SI.12}$$

SI.4. Mathematical formulary for system state analysis

Let C be the total number of countries c, T^* the set of hours $t^* \in [0..23]$ associated with the representative system states, $v_{c,t^*}^{orig,max}$ and $v_{c,t^*}^{orig,min}$ the maximum and minimum value occurring in the original data set for hour t^* for country c, v_{c,t^*}^{approx} the feature value

captured by the representative system state, respectively. The Mean Absolute Error of the system states to the corresponding mean (MAE_{AVG}), MAE_{MAX}, and MAE of the system states to the corresponding minimum (MAE_{MIN}) for one system element (load or a VRE generation) are then denoted as:

$$MAE_{AVG} = \frac{1}{C * T^*} \sum_{c \in C} \sum_{t^* \in T^*} |v_{c,t^*}^{orig,avg} - v_{c,t^*}^{approx}| \tag{SI.13}$$

$$MAE_{MAX} = \frac{1}{C * T^*} \sum_{c \in C} \sum_{t^* \in T^*} |v_{c,t^*}^{orig,max} - v_{c,t^*}^{approx}| \tag{SI.14}$$

$$MAE_{MIN} = \frac{1}{C * T^*} \sum_{c \in C} \sum_{t^* \in T^*} |v_{c,t^*}^{orig,min} - v_{c,t^*}^{approx}| \tag{SI.15}$$

SI.5. Implementation specifics and modifications to the model

The cluster analyses were conducted with the open-source environment Python. Wherever possible, predefined functionality is used, e.g. Pandas, NumPy, SciPy, and scikit-learn library [52–55]. If no other implementation is available, the algorithms are manually coded, e.g. the implementation for k-Medoids is based on [56], but enhanced to guarantee robustness and fit the purpose of this paper. The code is available upon request. Both, the k-Means++ implementation - which out-speeds k-Means in terms of convergence [44] - and the k-Medoids, employ Forgy's initialization [57]. Table SI.1 lists all Python packages and their version numbers used for the cluster analysis.

Table SI.1: Name and version number of Python packages used for clustering.

package	version	used functions for clustering
scipy	0.18.1	scipy.cluster.hierarchy: dendrogram, linkage, fcluster
		scipy.spatial.distance: pdist, euclidean, squareform
sklearn	0.0	sklearn.cluster: KMeans
sklearn_extensions	0.0.2	sklearn_extensions.fuzzy_kmeans: FuzzyKMeans

The ELTRAMOD-FBMC model is implemented in GAMS and draws upon the base model specified in [40]. Both the ATC mechanism and FBMC apply within the geographical scope of ELTRAMOD-FBMC. FBMC is implemented as in [58]. System states determined by a cluster analysis do not reflect diurnal structures of consecutive hours, but are selected based on similarity throughout the year. However, the storage functionality of pump storage power plants requires diurnal structures in the data. To overcome this

issue, endogenous storage variables are turned into exogenous parameters, which are fixed at the level of the endogenous model results based on the reference data set. Additional demand caused by pumping of pump storage power plants affiliated with a zone z is aggregated to a zone level and added to the derivation of the residual demand as exogenous parameter.

References

[1] M. Hoffmann, L. Kotzur, D. Stolten, M. Robinius, A review on time series aggregation methods for energy system models, Energies 13 (3) (2020) 641.

[2] P. Nahmmacher, E. Schmid, L. Hirth, B. Knopf, Carpe diem: A novel approach to select representative days for long-term power system modeling, Energy 112 (2016) 430–442. doi:10.1016/j.energy.2016.06.081.

[3] M. Robinius, A. Otto, P. Heuser, L. Welder, K. Syranidis, D. S. Ryberg, T. Grube, P. Markewitz, R. Peters, D. Stolten, Linking the power and transport sectors—part 1: The principle of sector coupling, Energies 10 (7) (2017) 956.

[4] L. Hirth, The market value of variable renewables. The effect of solar wind power variability on their relative price, Energy Economics 38 (2013) 218–236. doi:10.1016/j.eneco.2013.02.004.

[5] B. M. Reinhard, Complexity Reduction in Electricity Sector Modelling: A Comparative Analysis of Time-Frame Reduction Techniques, Master's thesis, Technical University Berlin (2015).

[6] C. Gerbaulet, F. Kunz, C. Lorenz, C. Von Hirschhausen, B. Reinhard, Cost-minimal investments into conventional generation capacities under a Europe-wide renewables policy, International Conference on the European Energy Market, EEM (2014) 1–7doi:10.1109/EEM.2014.6861297.

[7] K. Poncelet, H. Hoschle, E. Delarue, A. Virag, W. D'haeseleer, Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems, IEEE Transactions on Power Systems 32 (3) (2016) 1936–1948. doi:10.1109/TPWRS.2016.2596803.

[8] S. Pfenninger, Dealing with multiple decades of hourly wind and pv time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability, Applied energy 197 (2017) 1–13.

[9] G. Blanford, J. Merrick, D. Young, A Clean Energy Standard Analysis with the US-REGEN Model, The Energy Journal 35 (1) (2014) 137–164. doi:10.5547/01956574.35.SI1.8.

[10] F. De Sisternes, M. Webster, Optimal Selection of Sample Weeks for Approximating the Net Load in Generation Planning Problems, Massachusetts Institute of Technology Engineering Systems Division Working Paper Series: Citeseer (2013) 1–12.

[11] F. De Sisternes, M. Webster, I. Pérez-Arriaga, The Impact of Bidding Rules on Electricity Markets with Intermittent Renewables, IEEE Transactions on Power Systems 30 (3) (2015) 1603–1613. doi:10.1109/TPWRS.2014.2355775.

[12] F. M. B. Hobbs, J. Ho, S. Kasina, An engineering-economic approach to transmission planning under market and regulatory uncertainties: WECC case study, IEEE Transactions on Power Systems 29 (1) (2014) 307–317. doi:10.1109/TPWRS.2013.2279654.

[13] F. Munoz, A. Mills, Endogenous Assessment of the Capacity Value of Solar PV in Generation Investment Planning Studies, IEEE Transactions on Sustainable Energy 6 (4) (2015) 1574–1585. doi:10.1109/TSTE.2015.2456019.

[14] A. v. B. Hobbs, The economics of planning electricity transmission to accommodate renewables: Using two-stage optimisation to evaluate flexibility and the cost of disregarding uncertainty, Energy Economics 34 (6) (2012) 2089–2101. doi:10.1016/j.eneco.2012.02.015.

[15] G. Haydt, V. Leal, A. Pina, C. Silva, The relevance of the energy resource dynamics in the mid/long-term energy planning models, Renewable Energy 36 (11) (2011) 3068–3074. doi:10.1016/j.renene.2011.03.028.

[16] R. Kannan, H. Turton, A Long-Term Electricity Dispatch Model with the TIMES Framework, Environmental Modeling and Assessment 18 (3) (2013) 325–343. doi:10.1007/s10666-012-9346-y.

[17] S. Ludig, M. Haller, E. Schmid, N. Bauer, Fluctuating renewables in a long-term climate change mitigation strategy, Energy 36 (11) (2011) 6674–6685. doi:10.1016/j.energy.2011.08.021.

[18] M. Nicolosi, A. Mills, R. Wiser, The Importance of High Temporal Resolution in Modeling Renewable Energy Penetration Scenarios 1 Introduction and Motivation, 9th- Conference on Applied Infrastructure Research.

[19] A. Pina, C. Silva, P. Ferrão, Modeling hourly electricity dynamics for policy making in long-term scenarios, Energy Policy 39 (9) (2011) 4692–4702. doi:10.1016/j.enpol.2011.06.062.

[20] M. Welsch, P. Deane, M. Howells, B. O Gallachóir, F. Rogan, M. Bazilian, H. Rogner, Incorporating flexibility requirements into long-term energy system models - A case study on high levels of renewable electricity penetration in Ireland, Applied Energy 135 (2014) 600–615. doi:10.1016/j.apenergy.2014.08.072.

[21] J. Després, S. Mima, A. Kitous, P. Criqui, N. Hadjsaid, I. Noirot, Storage as a flexibility option in power systems with high shares of variable renewable energy sources: A POLES-based analysis, Energy Economics 67 (2017) 638–650. doi:10.1016/j.eneco.2016.03.006.

[22] J. Merrick, On representation of temporal variability in electricity capacity planning models, Energy Economics 59 (2016) 261–274. doi:10.1016/j.eneco.2016.08.001.

[23] L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, Impact of different time series aggregation methods on optimal energy system design, Renewable Energy 117 (2018) 474–487.

[24] H. Teichgraeber, A. R. Brandt, Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison, Applied Energy 239 (2019) 1283–1293.

[25] S. Pineda, J. M. Morales, Chronological time-period clustering for optimal capacity expansion planning with storage, IEEE Transactions on Power Systems 33 (6) (2018) 7162–7170.

[26] P. Härtel, M. Kristiansen, M. Korpås, Assessing the impact of sampling and clustering techniques on offshore grid expansion planning, Energy Procedia 137 (2017) 152 –

161, 14th Deep Sea Offshore Wind R&D Conference, EERA DeepWind 2017. doi:https://doi.org/10.1016/j.egypro.2017.10.342.

[27] S. Agapoff, C. Pache, P. Panciatici, S. Lumbreras, L. Warland, Snapshot selection based on statistical clustering for Transmission Expansion Planning, in: 2015 IEEE Eindhoven PowerTech, IEEE, 2015, pp. 1–6. doi:10.1109/PTC.2015.7232393.

[28] D. Fitiwi, F. De Cuadra, L. Olmos, M. Rivier, A new approach of clustering operational states for power network expansion planning problems dealing with RES (renewable energy source) generation operational variability and uncertainty, Energy 90 (2) (2015) 1360–1376. doi:10.1016/j.energy.2015.06.078.

[29] S. Fazlollahi, S. Bungener, P. Mandel, G. Becker, F. Maréchal, Multi-objectives, multi-period optimization of district energy systems: I. Selection of typical operating periods, Computers and Chemical Engineering 65 (2014) 54–66. doi:10.1016/j.compchemeng.2014.03.005.

[30] R. Green, I. Staffell, N. Vasilakos, Divide and Conquer? k-means clustering of demand data allows rapid and accurate simulations of the British electricity system, IEEE Transactions on Engineering Management 61 (2) (2014) 251–260. doi:10.1109/TEM.2013.2284386.

[31] T. Schütz, M. H. Schraven, M. Fuchs, P. Remmen, D. Müller, Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis, Renewable energy 129 (2018) 570–582.

[32] L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, Time series aggregation for energy system design: Modeling seasonal storage, Applied energy 213 (2018) 123–135.

[33] I. J. Scott, P. M. Carvalho, A. Botterud, C. A. Silva, Clustering representative days for power systems generation expansion planning: Capturing the effects of variable renewables and energy storage, Applied Energy 253 (2019) 113603.

[34] Y. Li, C. Liu, L. Zhang, B. Sun, A partition optimization design method for a regional integrated energy system based on a clustering algorithm, Energy 219 (2021) 119562.

[35] M. Zatti, M. Gabba, M. Freschini, M. Rossi, A. Gambarotta, M. Morini, E. Martelli, k-milp: A novel clustering approach to select typical and extreme days for multi-energy systems design optimization, Energy 181 (2019) 1051–1063.

[36] K. Kazor, A. S. Hering, Assessing the performance of model-based clustering methods in multivariate time series with application to identifying regional wind regimes, Journal of Agricultural, Biological, and Environmental Statistics 20 (2) (2015) 192–217.

[37] L. L. Tupper, D. S. Matteson, C. L. Anderson, L. Zephyr, Band depth clustering for nonstationary time series and wind speed behavior, Technometrics 60 (2) (2018) 245–254.

[38] F. Iglesias, W. Kastner, Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns, Energies 6 (2) (2013) 579–597. doi:10.3390/en6020579.

[39] A. Jain, R. Dubes, Algorithms for clustering data, Prentice Hall, Englewood Cliffs, NJ, 1988.

[40] D. Schubert, Bewertung von Szenarien für Energiesysteme - Potentiale, Grenzen und Akzeptanz, Ph.D. thesis, Dresden, Technical University (2016).

[41] T. Ladwig, Demand side management in deutschland zur systemintegration erneuerbarer energien.

[42] D. Schönheit, D. Hladik, H. Hobbie, D. Möst, Elmod documentation: Modeling of flow-based market coupling and congestion management, Tech. rep., Working paper of the Chair of Energy Economics (TU Dresden) (2020).

[43] A. Jain, M. Murty, P. Flynn, Data clustering: a review, ACM Computing Surveys 31 (1999) 264–323. doi:10.1145/331499.331504.

[44] D. Arthur, S. Vassilvitskii, 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, 2007, pp. 1027–1035.

[45] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, 3rd Edition, Elsevier Science, Amsterdam, 2012.

[46] A. Jain, Data clustering: 50 years beyond K-means, Pattern Recognition Letters 31 (2010) 651–666. doi:10.1016/j.patrec.2009.09.011.

[47] T. Warren Liao, Clustering of time series data - A survey, Pattern Recognition 38 (2005) 1857–1874. doi:10.1016/j.patcog.2005.01.025.

[48] G. Lance, W. Williams, A general theory of classificatory sorting strategies: 1. Hierarchical systems, The computer journal 9 (4) (1967) 373–380. doi:10.1093/comjnl/9.4.373.

[49] W. Day, H. Edelsbrunner, Efficient algorithms for agglomerative hierarchical clustering methods, Journal of Classification 1 (1984) 7–24. doi:10.1007/BF01890115.

[50] J. Ward Jr., Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association 58 (1963) 7–24. doi:10.1080/01621459.1963.10500845.

[51] C. Marton, A. Elkamel, T. Duever, An order-specific clustering algorithm for the determination of representative demand curves, Computers and Chemical Engineering 32 (6) (2008) 1373–1380. doi:10.1016/j.compchemeng.2007.06.010.

[52] W. McKinney, Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference (2010) 51–56.

[53] S. van der Walt, S. Colbert, G. Varoquaux, The NumPy Array: A Structure for Efficient Numerical Computation, Computing in Science & Engineering 13 (2) (2011) 22–30. doi:10.1109/MCSE.2011.37.

[54] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods 17 (2020) 261–272. doi:https://doi.org/10.1038/s41592-019-0686-2.

[55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, et al., Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[56] C. Bauckhage, NumPy / SciPy Recipes for Data Science : k -Medoids Clustering (2015). doi:10.13140/2.1.4453.2009.

[57] E. Forgy, Cluster analysis of multivariate data: efficiency versus interpretability models, Biometrics 61 (3) (1965) 768–769.

[58] K. Van den Bergh, J. Boury, E. Delarue, The Flow-Based Market Coupling in Central Western Europe: Concepts and definitions, Electricity Journal (1) (2016) 24–29. doi:10.1016/j.tej.2015.12.004.