Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 137 (2017) 152–161

Energy Procedia

www.elsevier.com/locate/procedia

14th Deep Sea Offshore Wind R&D Conference, EERA DeepWind'2017, 18-20 January 2017, Trondheim, Norway

Assessing the impact of sampling and clustering techniques on offshore grid expansion planning

Philipp Härtel[a,*], Martin Kristiansen[b], Magnus Korpås[b]

[a]*Energy Economy and Grid Operation, Fraunhofer IWES, Königstor 59, 34119 Kassel, Germany*
[b]*Department of Electric Power Engineering, NTNU, O. S. Bragstads plass 2E, 7034 Trondheim, Norway*

Abstract

Due to the ongoing large-scale connection of non-dispatchable renewable energy sources to the power systems, short- to long-term planning models are challenged by an increasing level of variability and uncertainty. A key contribution of this article is to explore and assess the implications of different dimension reduction approaches for long-term Transmission Expansion Planning (TEP) models. For the purpose of this study, a selection of sampling and clustering techniques are introduced to compare the resulting sample errors with a variety of sampling sizes and two different scaling options of the original data set. Based on the generated samples, a range of TEP model runs are carried out to investigate their impacts on investment strategies and market operation in a case study reflecting offshore grid expansion in the North Sea region for a 2030 scenario. The evaluations show that dimension reduction techniques performing well in the sampling and clustering process do not necessarily produce reliable results in the large-scale TEP model. Future work should include ways of incorporating inter-temporal constraints to better capture medium-term dynamics and the operational flexibility in power system models.

Peer-review under responsibility of SINTEF Energi AS.

Keywords:
Transmission Expansion Planning, Sampling, Clustering, Dimension Reduction, Offshore grids

1. Introduction

1.1. Increasing variability and uncertainty in TEP models

Most power systems around the world experience an increasing share of variable and non-dispatchable generation in their energy mix. At the same time, adequate models for both short-term and long-term planning become more complex. In comparison to traditional power systems which were primarily subject to power demand variations and fault occurrences, introducing high shares of renewable sources yields a significant rise of the power systems' underlying variability and uncertainty [1].

* Corresponding author. Tel.: +49 561 7249-471; fax: +49 561 7249-260.
E-mail address: philipp.haertel@iwes.fraunhofer.de

1876-6102
Peer-review under responsibility of SINTEF Energi AS.
10.1016/j.egypro.2017.10.342

Determining investments in new transmission lines or reinforcements of the existing transmission network is a crucial task in power system planning. These investment decisions are lumpy and capital intensive, which can have a long-lasting effect on expected market prices and power system operations. For this reason, the task of making sensible Transmission Expansion Planning (TEP) decisions is a widely studied problem [2].

The TEP is particularly relevant in the European context, where the European Union is pursuing a fully integrated internal energy market in which energy can flow freely across its regions. Robust transmission and distribution infrastructure, as well as a well-interconnected European network are seen as key constituents for a successful integration of renewable energy [3]. To be more specific, spatial levelling effects of fluctuating renewable energy resources, such as on- and offshore wind as well as solar, make grid reinforcements attractive [4]. With that in mind, recent developments, such as the aforementioned rise in variability and uncertainty, make efficient solutions of long-term TEP problems even more relevant, but at the same time increase their complexity.

1.2. Model complexity and computational challenges

In order to keep long-term TEP models tractable for a large geographical scope and a high level of spatial and temporal detail, a common approach is to use load duration curves or other generic scenario reduction approaches, such as sampling and clustering methods on the model's input data [5], [6], and [7]. For instance, a reduction approach focused on the model's output data rather than the input data is shown in [8]. Computationally, condensing the input data yields a smaller number of variables and constraints in the resulting optimization problems and leads to more acceptable solution times.

Dimension reduction can be crucial when dealing with large-scale planning models, as they often cover a multi-regional and multi-national scope. Given the broad geographical extent, location-specific climate- and weather-dependent characteristics cannot be omitted, as temperature, wind speeds and solar irradiation exhibit significant variations within the considered scope. Hence, it is of great importance to sustain the characteristic correlations when approximating full year time series with reduced-size, sampled time series. This is particularly valid for TEP, as the incentives for grid investments are triggered by spatial differentials, e.g. a high non-dispatchable production in one area with low demand could use a transmission line to transmit power to another area with high demand and low non-dispatchable generation.

1.3. Literature review

Regarding different dimension reduction approaches, a comprehensive and consistent comparison including a variety of sampling as well as clustering techniques is still not available from the literature. In [9], a number of partitioning and hierarchical clustering approaches are compared for probabilistic load modelling. Recently, a comparison of different approaches for selecting representative days in generation expansion planning problems as well as a new optimization-based approach is presented in [10]. Other works such as [11] present a comparison of different clustering techniques in the context of power system reliability assessments.

What is not yet clear is the impact of different dimension reduction methods on the results of TEP problems, such as the model for offshore grid expansion shown in this study. Metrics describing the quality of a raw data sample might significantly deviate from the effect it eventually has on the TEP model's quality of results which needs to be addressed. Therefore, it is the key objective of this study to assess the impact of different sampling and clustering techniques reducing the number of hourly time steps being considered by a long-term TEP model on its performance and the quality of its results.

In the remaining part of this article, Section 2 discusses the methodology used to carry out the comparative analysis of dimension reduction techniques and their consequences for a long-term TEP model. Section 3 provides an overview of the employed dimension reduction techniques in this study, i.e. sampling and clustering methods, and elaborates on the two scaling options applied in this article. Introducing the second phase of the study, Section 4 highlights the mathematical formulation of the long-term TEP model and the analysed case study reflecting an offshore grid expansion in the North Sea area. The first part of Section 5 presents the sampling results, and the second part exhibits the long-term TEP model results capturing the model-dependent effects of the dimension reduction techniques. In Section 6, the obtained comparison and evaluation results are discussed, and Section 7 concludes the study.

2. Methodology

The study was conducted in two phases. In the first phase, a comparative study of dimension reduction methods was performed. To this end, selected sampling and clustering techniques are introduced. These techniques are then used to sample from the full year time series data of a reference data set, which includes load, on- and offshore wind, solar, and hydro availability data. By using a variety of sample sizes and two different scaling options, the sampled time series data is compared against the reference data set to assess the techniques' respective impacts on the time series data with a reduced dimension.

In the second phase of the analysis, the sampled data of the first phase is used as input for a long-term TEP model. For this purpose, a deterministic TEP model is formulated and solved by mixed-integer linear programming (MILP). It co-optimizes investment decisions and market operation in a power system consisting of several market areas, Norway (NO), Great Britain (GB), Denmark (DK), Belgium (BE), Germany (DE), and the Netherlands (NL). Beyond that, the model is capable of optimizing combined HVAC and HVDC grids, with the latter being able to adopt both radial- and meshed structures. In terms of power electronics, meshed (multi-terminal HVDC) structures are the most advanced solution, but according to previous research also the most cost-effective alternative from a socio-economic point of view, see [12], for instance.

The case study reflects an offshore grid in the North Sea area with a scenario horizon of 2030 showing high shares of non-dispatchable power production capacity, predominantly solar and wind. It is based on the 'Vision 4' which was developed as one of the four contrasting visions for the long-term horizon 2030 by the European Network of Transmission System Operators for Electricity (ENTSO-E). These visions differ in terms of energy governance and ambitions towards the ongoing deployment of renewable energy sources [13].

Note that there are a few underlying assumptions which considerably help in maintaining a clear focus during the analysis. These assumptions are substantiated in Section 3 and their impact is reflected upon in Section 6.

3. Dimension reduction techniques

Given the main purpose of deriving a reduced representation of the time-dependent full year input data set, different sampling and clustering techniques are presented in this section. Clustering techniques divide a given set of data points into groups or clusters with the intention of having the data points belonging to one group to be more similar to each other than to the data points outside of the cluster. The full data set, that this section refers to, consists of a data matrix containing all relevant time series categories (e.g. hourly electricity load, wind power or solar feed-in in each market area) as columns, while the rows represent observations (hourly values of a full consecutive year).

It has to be stressed that the following comparative analysis is based on the premise that inter-temporal constraints, e.g. storage continuity equations of hydro reservoirs, are not explicitly considered in the second phase. This allows for an easier sampling of the input data since the chronological order of occurrence can be omitted. Given the method-oriented nature of this study, this approximation is considered to be reasonable. For a different approach incorporating transitions from one hourly system state to the other through the year, see [14].

When dealing with multivariate time series analysis, it is important to recognize the need for preparing the time series data in an adequate way. Because the relationship between all time-dependent data points plays a vital role in the sampling phase, the necessary first step of scaling or normalizing the input data needs to be addressed. Depending on the focus of the model, there exist different ways of preparing the time series data for the sampling and clustering process, as also described in [7]. Hence, the following two scaling options have been included in the analysis:

1. *Technology-specific scaling* by the highest occurring value across all market areas for load, onshore wind, offshore wind, solar, and hydro, respectively.
2. *Scaling by the highest occurring value* across all market areas in the full data set (peak-load in market area DE).

Scaling option 1 ensures that the maximum value of each technology type, e.g. wind offshore, is scaled to 1 across all market areas. As a result, the variability of each technology type is evenly weighed in the data set which is sampled from. By contrast, scaling option 2 yields a better representation of the aggregated power system, as it attributes its weighing according to the highest power consumption or generation values, albeit at the cost of e.g. a small country's representation exhibiting lower installed capacities.

Moreover, a heuristic based on a moving average is included as a further variant in the analysis. It is motivated by the fact that the clustering methods show a tendency to not capture the outliers so well, which is why an additional step was added to the sampling process. For each time series profile and all sampled points or clusters, data point values belonging to this sub-sample or cluster are compared to the moving average (6 h window) of the full year time series profile. If more than 95 % of these data point values are below or above the sampled or cluster value, the lowest or highest value within this sub-sample or cluster is chosen as the new sampled point or cluster center, respectively.

3.1. Systematic sampling

In comparison with the subsequent techniques, the systematic sampling is a simple approach. Hence, it can be regarded as a rather straight-forward but efficient method of producing the required samples. With this technique, elements are selected from an ordered sampling frame assuming that each element in the full data set has the same probability of being chosen (equiprobability). It starts by choosing an initial element from the time series data set and then selecting every k^{th} element, where the sampling interval k is determined by the desired sample size and the number of observations (8760 h of a full year in this case). As a minor improvement, the initial element, and its thereby determined sample, is not selected at random but rather phase shifted to the next second k^{th} element amounting to a set of k resulting samples. Out of these, the one showing the smallest average normalized root-mean-square error (NRMSE) of the full year reference is chosen as the final sample.

3.2. k-means clustering

The k-means clustering approach is a common technique using the k-means algorithm [15] (or Lloyd's algorithm [16]). It is an iterative, data-partitioning algorithm assigning n observations to exactly one of k clusters defined by centroids. Through this process, subsets of the full data set are created and the centroid of each subset corresponds to the mean of all measurements belonging to it. The sample size k needs to be chosen before the algorithm starts. In [6], k-means clustering is used for determining system states of wind and load data in a power system with high renewable penetration, for instance.

3.3. k-medoids clustering

The k-medoids clustering technique is similar to k-means as both partitioning methods try to divide a set of measurements or observations into k subsets so that the subsets minimize the sum of distances between a measurement and a center of the measurement's cluster. How the center or cluster of the subset is determined is the key difference between the k-means and k-medoids method. In the k-medoid algorithm, the center of the subset is an actual member of the data subset, called a medoid.

3.4. Hierarchical clustering

As a further dimension reduction method, the agglomerative form of hierarchical clustering analysis is used in this study. At each step, its underlying stepwise algorithm merges two objects, the ones with the least dissimilarity, thereby clustering the objects of the original data set. There are different ways of how the dissimilarities between clusters of objects or the linkage can be defined. Here, Ward's linkage [17] is used resulting in clusters with minimum inner squared distances (minimum variance algorithm). Hierarchical clustering with Ward's algorithm has been used in [7], for example, for the purpose of grouping similar days of full year data to decrease the dimension of a long-term power system model.

3.5. Moment-matching

In this context, the moment-matching technique refers to the approach presented in [18]. It belongs to a group of approaches aiming to minimize the selection of samples with respect to a predetermined criterion (external validity indices) [10]. This sampling algorithm selects a sample of hours from the full data set that minimizes the sum of square deviations of the moments between sampled hours and the full year time series data. The moments represent statistical measures such as correlation, mean and standard deviation. First, candidate samples are created by drawing 10,000 random samples from the full year data set. In order to then find appropriate estimators for the original time series data, the sample with the smallest squared moment deviation is chosen from the candidate samples. For instance, this technique has been successfully used for expansion planning with multivariate time series in [19].

4. Long-term TEP Model

This section gives a brief introduction to the long-term TEP model used in the second phase of the comparative analysis of dimension reduction techniques in this article. For more background information, see e.g. [20], [21], and [22] serving as a foundation for the following model (PowerGIM).

In order to incorporate uncertainty, the model is formulated as a two-stage stochastic program which relates to a mixed-integer linear program (MILP) in its extensive form. Integer variables are used to decide upon transmission infrastructure investments in the first stage, while the second stage problem is a pure linear program (LP) reflecting generator capacity investment and market operation. By only considering one scenario, the model is equivalent to a deterministic program. A compact model formulation of the stochastic MILP is given in (1) below.

$$TC = \min_x c^T x + E_\xi[\min_{y(\omega)} q(\omega)^T y(\omega)] \tag{1a}$$

s.t.

$$Ax \leq b \tag{1b}$$

$$T(\omega)x + Wy(\omega) \leq h(\omega), \quad \forall \omega \in \Omega \tag{1c}$$

$$x = (x_1, x_2) \geq 0, x_1 \in \{0,1\}, x_2 \in \mathbb{Z}^+, y(\omega) = (y_1(\omega), y_2(\omega)) \geq 0, \quad \forall \omega \in \Omega$$

In (1a), the objective function is divided into two stages; first the costs related to infrastructure investments, and second, the expected costs of market operation, $y_1(\omega)$, and generator capacity investments, $y_2(\omega)$, dependent on a discrete set of scenarios, Ω. For the work presented in this paper, generator capacity investments are disregarded in order to narrow down the scope to grid investments.

The vectors and matrices c, b, and A are associated with the first stage variables, i.e. investment in grid infrastructure. c is the cost vector for both fixed and variable node- and branch costs. b restricts the investment decisions, e.g. by the maximum allowed capacity per investment block (e.g. 1000 MW per branch), and A is the corresponding coefficient matrix to those investment constraints.

The second stage parameters are dependent on the realization of $\omega \in \Omega$, i.e. the parameters are not quantified before uncertainty is revealed. $q(\omega)$ is a cost vector for the marginal cost of generation and the capital capacity costs for generation. $h(\omega)$ is the right-hand-side restrictions for scenario ω, i.e. relevant restrictions on market dispatch and investments in generator capacity. $T(\omega)$ is the so-called *transition matrix* associated with first stage investments and it contains scenario and/or time-dependent data affecting operation in the second stage. The recourse matrix, W, is considered fixed in this model since as the coefficients in the matrix are independent of the realization of ω.

As stated earlier, only one scenario represented by ENTSO-E's Vision 4 is considered in the context of this study, that is the resulting model yields the same results as a deterministic program. Moreover, investment decisions are static implying that these are only made for one time step. Note that construction delays of investments are not considered. The economic lifetime of investments is assumed to be 30 years and the discount rate is 5 %. A CO_2 price amounting to 30 €/tCO_2 is used in order to reflect the social marginal cost of emissions from power plants based on fossil fuel, such as oil, gas, and coal.

5. Results

5.1. Sampling and clustering

The effect of using the two different scaling options is illustrated in Fig. 1. Because the load in market area DE contains the highest occurring value across all market areas, scaling option 2 results in a closer fit of the reference load profile than scaling option 1. By contrast, it can be seen that scaling option 1 produces a better match for the offshore wind profile. These observations correspond with the scaling methodology presented in Section 3.

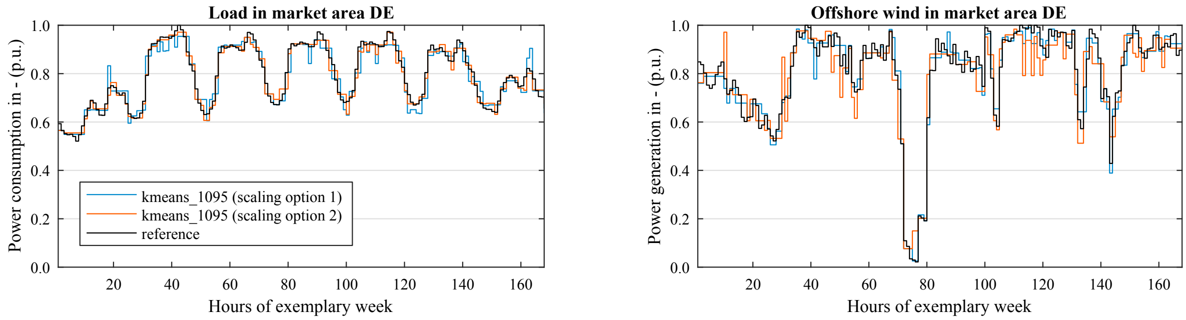

Fig. 1. Comparison of *k*-means clustering with two scaling options versus reference for load and offshore wind time series data in market area DE

The load level is the underlying driver for the resulting operational costs calculated by the TEP model. Fig. 2 provides useful information about the relative load levels resulting from the different sampling and clustering techniques.

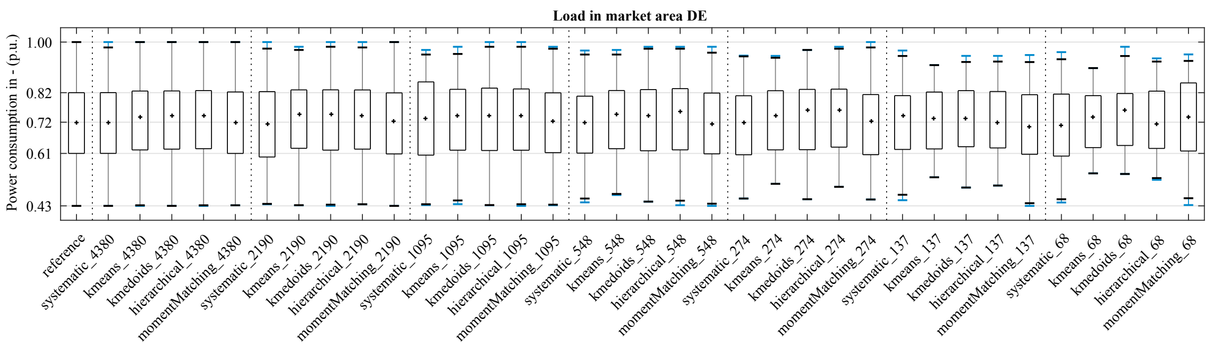

Fig. 2. Quartiles of reference and sampled data distributions for the load in market area DE (scaling option 1, effects of heuristic shown in blue)

For almost all techniques, the box plots suggest that the average load levels tend to be higher than in the reference case, although the highest values are not captured anymore. One exception, however, are the samples generated by the moment-matching method. Another important insight is that the heuristic introduced in Section 3 can partly capture the most extreme values of the original data. That said, the heuristic works better for the bigger sample sizes since it becomes harder to fulfill the 95 % criterion. In general, it must be noted that the heuristic comes at a cost, particularly for the clustering techniques, as its result deviates from the techniques' output (Fig. 3).

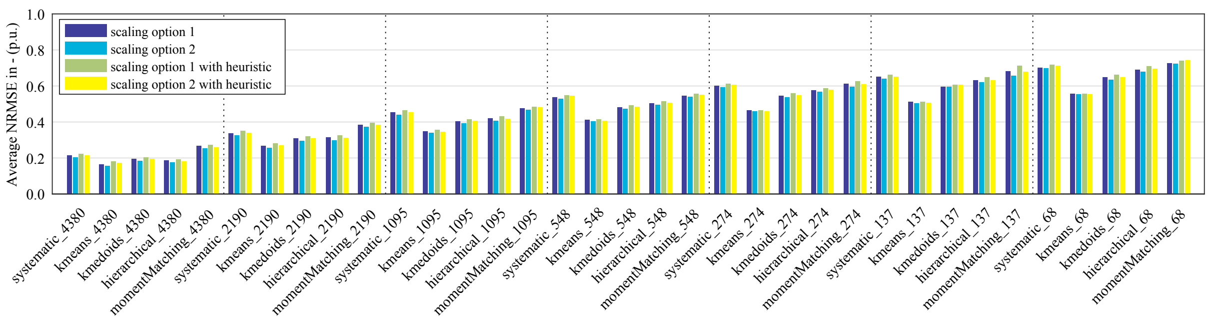

Fig. 3. Average normalized root-mean-square error (NRMSE) calculated over all time series categories for each technique and sample size

In order to quantify the overall fit in terms of profile deviations from the reference case, the NRMSE is calculated as an average for all time series, i.e. load, onshore wind, offshore wind, solar, and hydro, for each technique and sample size, as shown in Fig. 3. The NRMSE measure suggests that the *k*-means clustering performs best for all sample sizes, particularly with scaling option 2 without the heuristic algorithm (light blue bars). To put it another way, it stands to reason that k-means also yields the most accurate long-term TEP model results, which will be further assessed in the next section.

5.2. Long-term TEP case study

Based on the sampling and clustering results of Subsection 5.1, Table 1 gives an overview of the resulting key metrics of the long-term TEP case study simulations. As expected, with decreasing sample size the average solution time can be significantly reduced. To distinguish between pure market and investment decision effects, the total costs defined in the objective function (1a) are broken down into both operational and investment costs. Note that the share of operational costs is significantly higher than that of the investment cost. Keeping in mind that the k-means clustering performed best among the sampling and clustering results, it becomes clear that this is, on average, not the case for the model-dependent results reported in Table 1. In fact, it exhibits a poor performance regarding the average deviation in investment strategy and performs only slightly better than the systematic sampling when considering total cost deviations. The hierarchical clustering shows the highest average cost accuracy, followed by k-medoids.

Table 1. Average solution time reduction and average cost accuracy for each technique with respect to the full year reference case.

	Average reduction in solution time per sample size Solution time as share of full year reference in %							Average cost accuracy Deviation of full year reference in %		
	4380	2190	1095	548	274	137	68	Total (objective)	Investment	Operation
Systematic	17.83	5.69	2.11	1.03	0.36	0.17	0.09	1.48	0.90	1.51
k-means	23.11	5.75	2.14	0.86	0.62	0.21	0.11	-1.46	-3.36	-1.34
k-medoids	21.23	6.94	2.26	1.05	0.46	0.25	0.09	0.70	-1.63	0.84
Hierarchical	20.52	6.74	2.33	1.16	0.44	0.16	0.09	0.67	-0.23	0.72
Moment-matching	23.47	5.67	2.40	0.83	0.40	0.20	0.10	1.35	2.32	1.29
Reference (abs.)	——— 2016.1 s ———							473.1 bn€	26.9 bn€	446.1 bn€

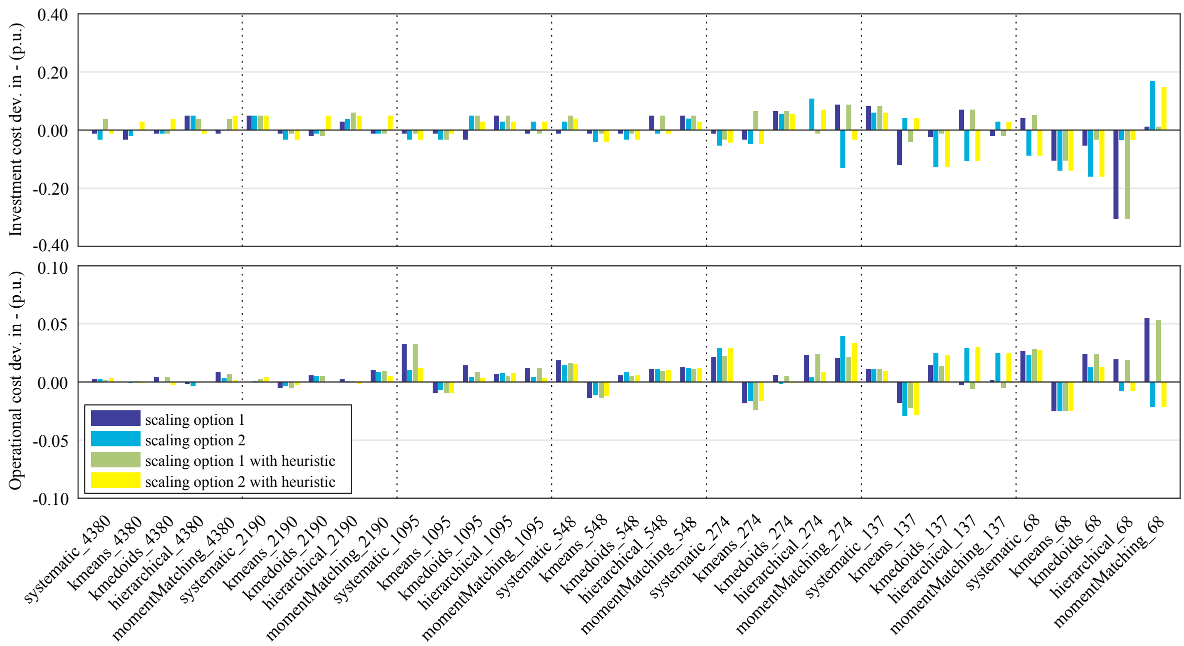

Fig. 4. Comparison of relative investment and operational cost deviations or each technique and sample size (sampled - reference)

5.2.1. *Investment and operational costs*

In Fig. 4, a more detailed breakdown of the relative investment and operational cost deviations is presented. Several results can be taken from this figure: First, for all methods, both investment and operational cost deviations generally increase with a reduced sample size. Second, while the other methods rather overestimate the operational costs, *k*-means shows a consistent underestimation confirming the observation above. Third, the scaling options do seem to have a bigger impact on the deviations than applying the heuristic. More specifically, hierarchical clustering seems to work slightly better with scaling option 1, while particularly for *k*-medoids, systematic sampling, and moment-matching, scaling option 2 presents a better combination. However, there is no clear indication as to which scaling option performs better.

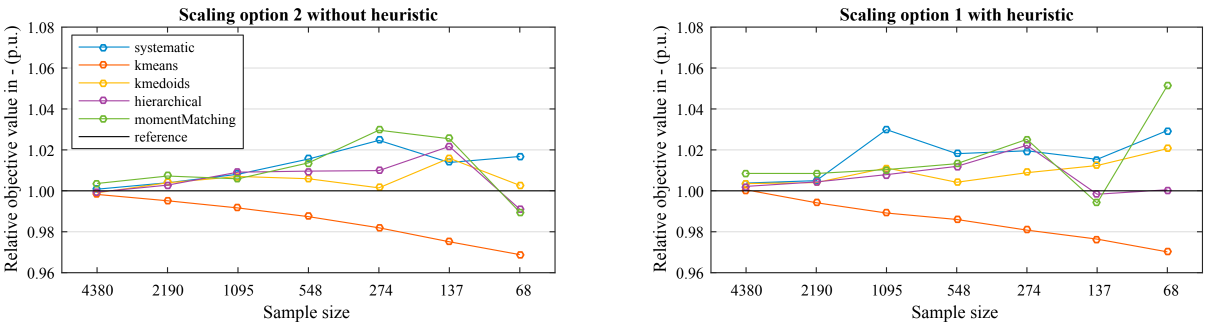

Fig. 5. Convergence of the relative objective value for each sample size and method, scaling option 2 without heuristic (left) and scaling option 1 with heuristic (right)

As can be seen from Fig. 5, the convergence results of the relative objective are in line with the previous findings. For the half year sample size (4380 h), all techniques show relative values close to 1. At the opposite end, the moment-matching technique displays a deviating behavior for the 68 h sample size. Interestingly, hierarchical clustering achieves relatively low deviations for the smallest sample size with scaling option 1 and the applied heuristic.

5.2.2. *DC cable investments*

In Fig. 6, the resulting differences in investment strategy are presented as deviations in DC cable investments.

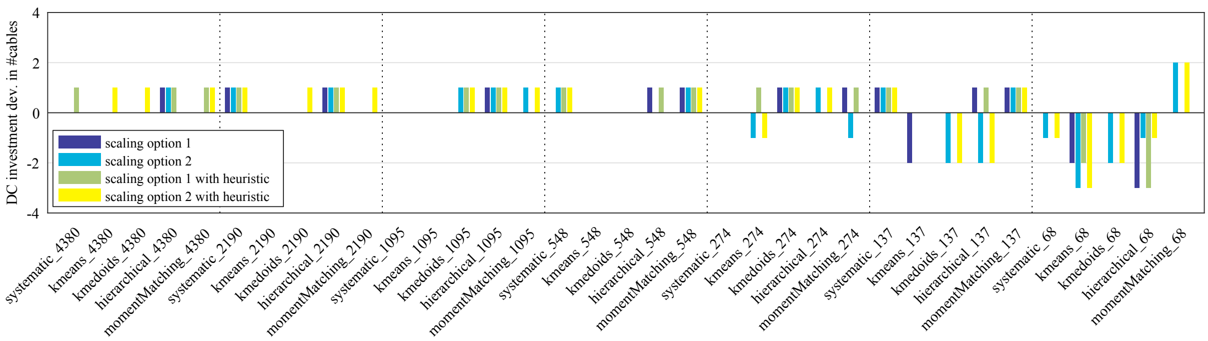

Fig. 6. Comparison of DC cable investment deviations or each technique and sample size (sampled - reference)

Out of the 14 DC cable investments in the full year reference, this finding implies that over-investments are limited to one DC cable, except for the moment-matching technique with the smallest sample size. Under-investments do not occur for sample sizes bigger than 274 h, while they can amount to three DC cables for the smaller sample sizes. Hence, there is reason to believe that the high load levels indicated in Fig. 2 might be the explanatory driver for the occurring over-investments. This is because the additional transmission capacity is used to cover the loads more efficiently with cheap generation technologies located elsewhere.

6. Discussion

There are a few assumptions and limitations to be kept in mind when discussing the results obtained in the comparisons above. As stated earlier, short- to long-term inter-temporal constraints (e.g. seasonal hydro reservoir continuity) are not accounted for in the TEP model. At the same time, however, this assumption facilitates the sampling and clustering of individual hours because the chronology of the original time series data can be ignored. Further, the number of considered market areas and technologies, and thus the number of time series correspond to a full-size problem. The effect of increasing the time series quantity of the original data set, e.g. from a two time series test case, has not been investigated. This fact should not be neglected, however, since a larger dimensionality of the original data set requires more observations to obtain reasonable sampling and clustering results.

The analysis suggests that the dimension reduction techniques may score very well in terms of capturing distributions or error measures such as the presented NRMSE, but that the output from the TEP model employing the samples gives deviating score patterns in terms of investment and operational costs of a potential future offshore grid. This insight becomes particularly obvious for the k-means clustering approach, which is performing well in Subsection 5.1 but consistently underestimating the total costs in the TEP model runs in Subsection 5.2. Comparing it with k-medoids, the way of determining the centroids seems to play an important role. The agglomerative hierarchical and k-medoids clustering show comparatively good results when quantifying both the NRMSE and the effects on offshore grid expansion decisions in the North Sea case study.

It has been shown that the scaling options have a greater impact than the applied heuristic. Then again, no clear indication can be given as to the more suitable choice of either one of the two scaling options. Hence, paying careful attention to different scaling options for the original data set seems appropriate.

7. Conclusion

Motivated by the concern of growing model complexity and increasing computational challenges, this article investigates the impact of dimension reduction methods for power system models. To this end, a selection of dimension reduction techniques is analysed and used to sample from hourly full year time series data including load and renewable generation.

The main contributions include a comprehensive comparison of sampling and clustering techniques with different scaling options which were used in combination with a large-scale TEP model for a North Sea offshore grid case study. Further, a high number of market areas and technology options, i.e. large number of time series categories, was considered in this study. It can be concluded that techniques performing well in the sampling and clustering process do not necessarily produce reliable results in the large-scale TEP model.

A subsequent analysis of dimension reduction techniques and their application in long-term power system models can include the use of more sophisticated heuristics, particularly in investment models as they significantly depend on the highest occurring values in the original data sets. Future work should include ways of incorporating inter-temporal constraints to better capture medium-term dynamics and the operational flexibility in power system models. For instance, this could be done by employing dimension reduction approaches, e.g. [14], or developing alternative solution strategies involving decomposition for the full year problem.

Acknowledgements

The work has been performed in the framework of the NSON Initiative. Funding has among others come from the German "North Sea Offshore Network" (NSON-DE) project, financed as part of the funding initiative "Zukunftsfähige Stromnetze" (FKZ 0325576H) by the German Federal Ministry for Economic Affairs and Energy (BMWi) and cooperation has been facilitated by the IRP Mobility Programme.

The long-term TEP model used in this article, PowerGIM, is an open-source model available at Bitbucket. The authors would like to thank Gurobi for a powerful optimization solver.

References

[1] Tahanan, M., van Ackooij, W., Frangioni, A., Lacalandra, F.. Large-scale unit commitment under uncertainty. 4OR 2015;13(2):115–171. doi:10.1007/s10288-014-0279-y.

[2] Lumbreras, S., Ramos, A.. The new challenges to transmission expansion planning. survey of recent practice and literature review. Electric Power Systems Research 2016;134:19–29. doi:10.1016/j.epsr.2015.10.013.

[3] European Commission. Clean Energy For All Europeans, COM(2016) 860 final. 2016. URL: http://ec.europa.eu/energy/en/news/commission-proposes-new-rules-consumer-centred-clean-energy-transition.

[4] European Commission. Connecting power markets to deliver security of supply, market integration and the large-scale uptake of renewables. 2015. URL: http://europa.eu/rapid/press-release_memo-15-4486_en.htm.

[5] Hobbs, B.F., Ji, Y.. Stochastic programming-based bounding of expected production costs for multiarea electric power system. Operations Research 1999;47(6):836–848. doi:10.1287/opre.47.6.836.

[6] Wogrin, S., Duenas, P., Delgadillo, A., Reneses, J.. A new approach to model load levels in electric power systems with high renewable penetration. IEEE Transactions on Power Systems 2014;29(5):2210–2218. doi:10.1109/TPWRS.2014.2300697.

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

[8] Fitiwi, D.Z., de Cuadra, F., Olmos, L., Rivier, M.. 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 2015;90:1360–1376. doi:10.1016/j.energy.2015.06.078.

[9] ElNozahy, M.S., Salama, M.M.A., Seethapathy, R.. A probabilistic load modelling approach using clustering algorithms. In: IEEE Power and Energy Society general meeting (PES), 2013. Piscataway, NJ; 2013, p. 1–5. doi:10.1109/PESMG.2013.6672073.

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

[11] Kile, H.. Evaluation and grouping of power market scenarios in security of electricity supply analysis. Doctoral thesis; Norwegian University of Science and Technology; Trondheim; 2014.

[12] Kristiansen, M., Korpås, M., Farahmand, H., Graabak, I., Härtel, P.. Introducing system flexibility to a multinational transmission expansion planning model. In: 2016 Power Systems Computation Conference (PSCC). 2016, p. 1–7. doi:10.1109/PSCC.2016.7540861.

[13] ENTSO-E. TYNDP 2016 Scenario Development Report. 2015. URL: http://www.entsoe.eu/.

[14] Wogrin, S., Galbally, D., Reneses, J.. Optimizing storage operations in medium- and long-term power system models. IEEE Transactions on Power Systems 2016;31(4):3129–3138. doi:10.1109/TPWRS.2015.2471099.

[15] MacQueen, J.B.. Some methods for classification and analysis of multivariate observations. In: Le Cam, L.M., Neyman, J., editors. Fifth Berkeley Symposium on Mathematical Statistics and Probability; vol. 1. Berkeley: University of California Press. ISBN 0097-0433; 1967, p. 281–297. URL: http://projecteuclid.org/ euclid.bsmsp/1200512992.

[16] Lloyd, S.. Least squares quantization in pcm. IEEE Transactions on Information Theory 1982;28(2):129–137. doi:10.1109/TIT.1982.1056489.

[17] Ward Jr., J.H.. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 1963;58(301):236–244. doi:10.1080/01621459.1963.10500845.

[18] van der Weijde, A.H., Hobbs, B.F.. Planning electricity transmission to accommodate renewables: Using two-stage programming to evaluate flexibility and the cost of disregarding uncertainty. 2011. doi:10.17863/CAM.5218.

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

[20] Trötscher, T., Korpås, M.. A framework to determine optimal offshore grid structures for wind power integration and power exchange: A framework to determine optimal offshore grid structures. Wind Energy 2011;14(8):977–992. doi:10.1002/we.461.

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

[22] Svendsen, H.G., Spro, O.C.. PowerGAMA: A new simplified modelling approach for analyses of large interconnected power systems, applied to a 2030 Western Mediterranean case study. Journal of Renewable and Sustainable Energy 2016;8(5):055501. doi:10.1063/1.4962415.