Energy Strategy Reviews 56 (2024) 101569

Contents lists available at ScienceDirect

Energy Strategy Reviews

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

Avoiding false inter-zonal meshing in the clustering of a large-scale German power grid

Katharina Esterl [a,*], Carlos Andrés Epia Realpe [b], Ulf Philipp Müller [b]

[a] Europa-Universität Flensburg, Auf dem Campus 1, Flensburg, 24943, Germany
[b] Hochschule Flensburg, Kanzleistraße 91-93, Flensburg, 24943, Germany

ARTICLE INFO

Dataset link: https://openenergy-platform.org/, https://zenodo.org/records/8376714, https://github.com/openego/eTraGo/tree/0.9.0, https://etrago.readthedocs.io/en/latest/index.html, https://github.com/openego/eTraGo/blob/paper/kmedoids-dijkstra-2/etrago/args.json, https://github.com/openego/eTraGo/blob/58b953fa3a2a51e30855d9ee73e9758b8699a2a6/etrago/cluster/spatial.py#L681, https://github.com/openego/eTraGo/blob/58b953fa3a2a51e30855d9ee73e9758b8699a2a6/etrago/cluster/electrical.py#L426

Keywords:
Energy system modelling
Power grid modelling
Power grid complexity reduction
k-means clustering
Open source
Open data

ABSTRACT

The ongoing transformation towards a sector-coupled energy system based on renewable generation leads to more complex grid-based energy system models. It is a crucial task to reduce the models' complexities in order to keep optimisation problems tractable while at the same time generating appropriate results. This work adds a *Dijkstra's algorithm* to a commonly used *k-means Clustering* method to reduce the spatial complexity of a German transmission and sub-transmission grid model. The novel approach leads to more accurate results while reaching faster calculation times. In particular, it successfully avoids false inter-zonal meshing. Consequently, the more accurate modelling of inter-cluster connections results in up to 41% higher grid expansion needs and significantly changed spatial allocations of network and storage expansion. Where geographical or political borders (e.g. the former inner-German border) have led to scarcely interconnected grid topologies (especially on the sub-transmission grid level), the impact is particularly high. The presented work follows open-source and open-data principles.

1. Introduction and state of the art

The transformation of the energy system is crucial to mitigate anthropogenic climate change. Modelling and optimisation of power grids are used to investigate the challenges and opportunities within the transformation process (cf. [1,2]). As a result of the increase of the share of renewable energies (RE) in electricity generation along with progressing sector coupling, current energy system models show increasing complexity. Therefore, complexity reduction methods are needed to optimise these systems within an acceptable computational time.

The complexity of energy system models depends among other things on the complexity of the input data, which is characterised by its spatial and temporal resolution [3]. In order to reduce complexity in those dimensions, methods are being developed, examined and finally applied. Spatial complexity is defined by the number of buses (nodes in a network) and their connections within the network. The aim of spatial complexity reduction methods is to reduce the number of buses and the number of connections while preserving relevant information about the original network in order to study the energy system, e.g. potential grid expansion needs. Aryanpur et al. [4] stress the advantages of considering high spatial resolutions when investigating on grid expansion needs. Yet, as Dashti and Rouhandeh [5] emphasise, the trade-off between the spatial complexity and computational limitations should be addressed.

A commonly used approach to reduce the spatial complexity is the *k-means Clustering* (cf. [6–13]). This method, explained in detail by Hörsch and Brown [6], implies a clustering of network buses into a predefined number of groups, taking into account the geographical positions of the buses. A *k-means Clustering* aims to minimise distances within a cluster and maximise distances between different clusters by usually considering the Euclidean distances between the buses. In most cases, the buses, and therefore also the distances, are weighted taking into account generation, load and storage capacities. After clustering, the resulting clusters are replaced by representative buses with certain

* Corresponding author.
E-mail address: katharina.esterl@uni-flensburg.de (K. Esterl).

https://doi.org/10.1016/j.esr.2024.101569
Received 29 July 2023; Received in revised form 2 July 2024; Accepted 21 October 2024
Available online 12 November 2024

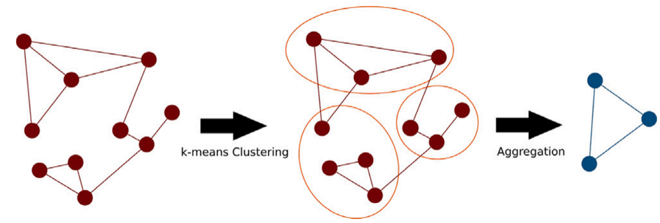


Fig. 1. False inter-zonal meshing within *k-means Clustering*.

properties that represent all buses per cluster at the geographical average per cluster. Connections between original buses in different clusters are aggregated to connections between representative buses. Thus, the original network with its original number of buses and connections is replaced by a network with a lower number of buses corresponding to the chosen number of clusters and a lower number of connections.

In the resulting reduced networks, buses are sometimes assigned to cluster centres that are geographically close (in terms of the Euclidean distance) but electrically far away (cf. [14,15]). This is particularly the case for stub lines and stub meshes (cf. Fig. 1). As it can be seen in Fig. 1, in the course of aggregation, lines are formed between clusters that have no direct electrical connection (cf. [15]). Therefore, false inter-zonal (i.e. inter-cluster) meshing occurs (cf. [14]). In these cases, the topology of the complexity-reduced network does not represent sufficiently the original network. Existing grid capacities tend to be overestimated and structural changes lead to incorrectly represented reactances resulting in a incorrectly modified power flow behaviour. Consequently, expansion needs are likely to be underestimated.

Methods have been developed that take into account the electrical topology and thus avoid the problem of false inter-zonal meshing. In this sense, Weber et al. [16] developed an interesting approach, but only applied it to a detailed, regional subsystem of the German power grid. Consequently, the degree of reduction is quite small and the applicability on a large-scale power grid system is not demonstrated. Biener and Garcia Rosas [15] apply an agglomerative clustering of buses using the electrical distance quantified by the reactance between buses as a distance measure. The aim is to cluster regions of high electrical connectivity to minimise errors in load flow behaviour. The paper by Frysztacki et al. [17] compares several approaches including the aforementioned method by Biener and Garcia Rosas [15] which is further developed by taking into account the admittance and therefore also including the resistance. Whereas the latter two mentioned approaches apply their clustering on particular large-scale models considering the European and German transmission grid (namely *PyPSA-Eur* and *ELMOD-DE*), the method introduced in the present work is applied to a model that additionally includes the sub-transmission grid (i.e. the HV grid at 110 kV) resulting in a system with significantly more buses. This makes the model of the entire German electric system uniquely detailed while taking into account neighbouring countries.

This paper introduces a novel, two-step method called *k-medoids Dijkstra Clustering* which in the first step is highly oriented on the original *k-means Clustering* method and in the second step adds a *Dijkstra's algorithm*. The *Dijkstra's algorithm* aims to identify the shortest path distances between buses. The path distances are defined by the accumulated line lengths. Hence, the network is reduced by taking its topology into account. Therefore, problems of false inter-zonal meshing are addressed while maintaining the weighting of the buses and the selection of cluster centres on the basis of the Euclidean distance. Moreover, pure line lengths are considered instead of electrical parameters. This provides a more available and therefore reliable source in the context of large-scale open data models.

The *k-medoids Dijkstra Clustering* has already been applied to a sector-coupled system without examining the method's impact on the optimisation results (cf. [18]). In contrast, the present study evaluates the impact of the novel method through a comparative study that directly addresses the clustering results as well as the subsequent optimisation results, in particular the spatial allocation of network and storage expansion. Thus, this work represents an in-depth analysis of the effects resulting from the application of the *k-medoids Dijkstra Clustering* in comparison to the use of a commonly used *k-means Clustering*, with special focus on the avoidance of false inter-zonal meshing in the German transmission and sub-transmission grid. Since the developed method focuses on the power grid, and in order to essentially evaluate interrelations and effects more easily, a reduced power-only model is used. The used grid model is characterised by a comparatively high spatial resolution. Particularly within the sub-transmission grid level, the topology shows a significant number of stub lines and stub meshes. Therefore, differences in the reduced networks are expected when applying the different considered methods. Additionally, as the *k-medoids Dijkstra Clustering* is assumed to lead to a more precise representation of the original topology, a more accurate prediction of expansion needs is expected.

Consequently, the following research questions will be answered:

(a) Does the usage of the *k-medoids Dijkstra Clustering* significantly avoid false inter-zonal meshing when clustering the German transmission and sub-transmission grid?
(b) How does the developed method affect the optimisation of capacities and spatial allocation of network and storage expansion?

2. Methods

2.1. General model setup

The *k-medoids Dijkstra Clustering* was implemented as part of the research project *eGo^n*.[1] In this project, a model was developed that covers several energy sectors and thus assesses the impact of sector coupling on both transmission and distribution electricity grids in Germany. In the present work, the optimisation tool *eTraGo* and the underlying data model, including in particular the German eHV and high voltage grid, are used. As the work follows open source and open data principles, all data and tools are publicly available (cf. Section Data availability).

The German energy system is defined in an *eGon2035*-scenario with total generation capacities for the year 2035 based on the scenario *C2035* defined by German Transmission System Operators [19]. For the neighbouring countries, similar information is taken from the *Ten-year network development plan (TYNDP) 2020 scenario 'distributed energy'*. The temporal complexity is characterised by 8,760 time steps (one year in hourly resolution). To address computational limitations, for the optimisation, this resolution is reduced to every fifth time step.

The topology of the used German power grid model includes the eHV and HV level (cf. Fig. 4(a)). Substations and lines are extracted from OpenStreetMap (OSM) [20] and electrically parameterised using the tool *osmTGmod* [21–23]. The original spatial complexity is defined by 9,345 buses and 10,308 lines. This high spatial resolution model has been used in various scientific works (e.g. [10,14,18]). Addressing computational burden in these previous works, the grid model was reduced to a common standard setting with 300 buses.

The tool *eTraGo* performs a Linear Optimal Power Flow (LOPF) using the python library *PyPSA* [24]. The objective of the optimisation, described in [24], is to minimise the annual system costs, defined as the sum of annualised network and storage expansion costs and annual dispatch costs (cf. Appendix A). More detailed information on the model, the scenario and the optimisation tool *eTraGo* including the used settings can be found in [18]. Within Section Data availability, a reference to the parameterisation can be found. In contrast to the model in [18], as already motivated in the previous section, a reduced power-only model is used. Thus, the heat, mobility and gas sectors are neglected.

[1] Project website: https://ego-n.org/.

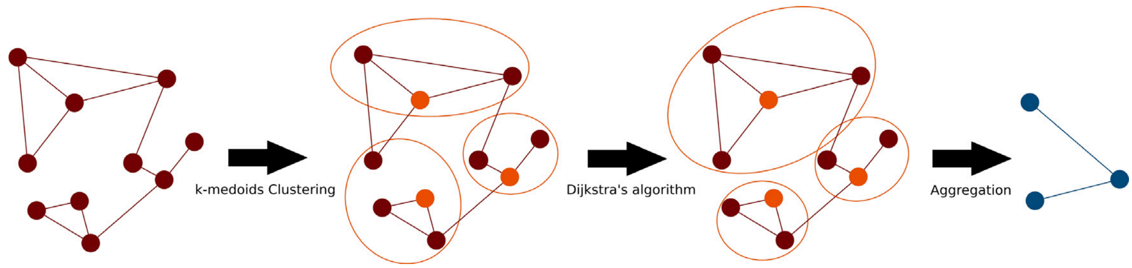


Fig. 2. Rearrangement of clusters by adding a *Dijkstra's algorithm*.

2.2. The k-medoids Dijkstra Clustering

In the first step, the developed clustering method implies a *k-medoids Clustering* (cf. Fig. 2). The aim of this step is to identify the set of buses to which all other buses are assigned by using a *Dijkstra's algorithm* in the second step. To be able to apply a *Dijkstra's algorithm*, a *k-medoids Clustering* is needed instead of a *k-means Clustering*, as the obtained buses need to be part of the original topology.

This clustering approach is very similar to the *k-means Clustering* described in Section 1 (cf. Fig. 1, explained in [6]). The selection of the medoid's position x_c for $c = 1, \dots k$ is defined by Eq. (1). Thus, apart from linearly weighting the original buses $n \in N$ with respect to the connected generation, load and storage capacities (w_n), the Euclidean distance between buses is used in order to find the clusters C. Instead of defining new positions for network buses at the geographical centroids of the clusters (as in the *k-means Clustering*), in the *k-medoids Clustering*, the nearest original bus of each centroid is used as the medoid of the cluster [25]. Consequently, each cluster c is represented at the position of an actual bus in the original network ($x_c \in X_n$) having a minimum Euclidean distance to the remaining buses of its cluster N_c.

$$\min_{x_c} \sum_{c=1}^{k} \sum_{n \in N_C} w_n \left\| x_c - x_n \right\|^2 \quad \text{with} \quad x_c \in X_n \tag{1}$$

In a second step, a *Dijkstra's algorithm* is performed (cf. Fig. 2). In this step, all buses within the original network are assigned to the previously identified set of medoid buses depending on their shortest paths to them.

In general, as described by Cormen et al. [26], the *Dijkstra's algorithm* identifies the shortest path between a starting vertex and one of the remaining vertices or all the remaining vertices in a weighted graph $G(V, E)$ with $w_e(u, v) \geq 0$ for each edge $(u, v) \in E$. In an initialisation step, the path distance to the start vertex is given the value 0 ($\delta(y) = 0$), and all the other path distances are defined as ∞ ($\delta(z) = \infty$). A set of vertices is defined containing all vertices already considered ($\sigma = (v_y)$). Now, all vertices are checked iteratively ($\sigma = (v_1, v_2 ... v_n) \forall v \in V$), each time calculating the shortest path distance $\delta(z)$ and updating the path $P(z)$ whenever the calculation results in a reduction of the path distance, as defined in Eq. (2).

$$\begin{aligned} &\delta(z) = min\{\delta(z), \delta(v_i) + w_e(v_i, z)\} \\ &P(z) := \begin{cases} P(y, v_i, z) & \text{if } \delta(z) < \delta(v_i) + w_e(v_i, z) \\ P(y, v_1, z) & \text{otherwise} \end{cases} \end{aligned} \tag{2}$$

Therefore, a graph $G(V, E)$ is built with graph vertices V representing the original network buses N and graph edges E representing the power lines weighted by their line length $w_e(u, v)$. Using the *Dijkstra's algorithm*, the shortest paths from each medoid to all buses of the original network are calculated, where the path distances are defined by the line lengths between them. Finally, the original buses are assigned to the medoid that has the shortest path distance to reach it. Consequently, as seen in Fig. 2, the assignment of the buses to the individual clusters (based on the *k-medoids Clustering*) is potentially changed when applying the *Dijkstra's algorithm*. Hereby, the grouping of electrically distant buses is avoided.

In order to aggregate the buses per cluster, the attributes of the representative buses are determined as done in the *k-means Clustering* and described in [6], while keeping the geographical positions of the medoids. Note that, as a prerequisite for the clustering and the subsequent aggregation, all buses and lines with different voltage levels have been converted to 380 kV by adapting the corresponding electrical parameters as defined in [10]. Whereas the aggregation of lines leads to a false inter-zonal meshing when applying a *k-means Clustering* (cf. Fig. 1), the overestimation of connections between different clusters is avoided by modified assignments, as it can be seen in Fig. 2. The effectiveness of the developed method has been proven by analysing minimal example network typologies, focusing in particular on the topological changes induced by the *Dijkstra's algorithm* and manually recalculating the shortest paths.

The method is implemented in the optimisation tool *eTraGo* (cf. Section 2.1), which is based on the *PyPSA* library [24]. The `kmedoids_dijkstra_clustering` function includes the described steps of the approach. First, the *k-medoids Clustering* is executed using the *scikit-learn* package [27]. Then, the function `dijkstras_algorithm` is called to perform the *Dijkstra's algorithm*, which is implemented using the package *NetworkX* [28]. Finally, the aggregation is performed, again utilising the *scikit-learn* package. The code is publicly available and documented (cf. Section Data availability).

2.3. Evaluation approach

The novel clustering method, namely the *k-medoids Dijkstra Clustering* (detailed in Section 2.2) is comparatively evaluated in Sections 3 and 4. The *k-medoids Dijkstra Clustering* and a *k-means Clustering* are applied to reduce the spatial complexity of the German transmission and sub-transmission grid (introduced in Section 2.1) to networks with resolutions in the range of 50 to 700 buses while the main focus remains on the standard 300-bus system.

First, the clustered grid topologies are compared to the original topology (see Section 3.1). In order to assess the quality of the clustered networks, their modularities as defined by Clauset et al. [29] and their electrically weighted modularities as defined by Biener and Garcia Rosas [15] and Frysztacki et al. [17] (for the mathematical description see Appendix B) are calculated ex post. Biener and Garcia Rosas [15] and Frysztacki et al. [17] have shown a positive correlation between the modularity of a clustered network and its adequacy. Using this correlation, they maximise modularity as an intrinsic evaluation measure to obtain the desired clustered networks. Since modularity generally reaches high values in clustered networks with many intra-cluster connections and few inter-cluster connections, the inter-cluster meshing degrees of the reduced networks are additionally evaluated. Furthermore, inter-cluster links crossing the former inner-German border are evaluated as a case study. In this way, specific effects of the new method can be demonstrated.

Secondly, the effects on the optimisation results are evaluated (see Section 3.2). In this context, the resulting system costs as well as grid and storage expansion requirements within Germany and their spatial allocation are compared with particular focus on the former inner-German border. Since the annual optimisations described cannot be

Table 1
Quantification of the degree of meshing and changes in assignment by adding a *Dijkstra's algorithm*.

Number of buses	Number of lines		*k-medoids Dijkstra Clustering*	
	k-medoids Dijkstra Clustering	*k-means Clustering*	Number of re-assigned original buses	Share of re-assigned original buses
50	67	80	1736	19%
100	170	212	2272	25%
300	533	669	2743	30%
500	833	1014	2662	29%
700	1090	1336	2523	28%

Table 2
Comparison of the transmission capacity across the former border between former FRG and GDR for the clustered networks in different spatial resolutions. The capacity in the original network amounts to a total of 19.2 GVA (17.9 GVA on eHV level and 1.3 GVA on HV level). For the 300-bus system, the original grid level of the representative lines is indicated.

Number of buses	Transmission capacity in reduced networks in GVA		Error compared to original transmission capacity (19.2 GVA)	
	k-medoids Dijkstra Clustering	*k-means Clustering*	*k-medoids Dijkstra Clustering*	*k-means Clustering*
50	39.5	47.7	106%	148%
100	19.5	35.0	1%	82%
300	19.5	40.4	1%	110%
- eHV level	17.9	27.1	0%	51%
- HV level	1.5	13.3	15%	923%
500	18.9	25.7	−1%	34%
700	18.9	28.5	−1%	48%

performed in full spatial complexity due to computational limitations, an extreme case, defined by the time step with the highest overall absolute line loading, is analysed to allow a comparison with the optimised system in full spatial complexity. Furthermore, sensitivity analyses were performed in order to evaluate the variance and robustness of the results.

3. Results

3.1. Clustering results

The newly developed *k-medoids Dijkstra Clustering* leads to significant topological changes in the reduced networks. As the *k-medoids Dijkstra Clustering* includes a *k-means Clustering*, it is possible to directly evaluate how many original network buses are reassigned when applying the *Dijkstra's algorithm*. For our standard case, i.e in the reduced system with 300 buses, almost 30% of the original buses are reassigned (cf. Table 1). For systems with higher spatial resolutions the reassignment rate remains fairly constant, although slightly decreasing, while for lower resolutions, it decreases more significantly. Consequently, the reassignment of original buses leads to fewer inter-cluster connections, i.e. fewer lines in the reduced networks (cf. Table 1). The degree of meshing in the *k-medoids Dijkstra* reduced networks is therefore significantly lower. In the case of a reduced system with 300 buses, the number of inter-cluster lines decreases by 20%. A similar relative difference can be depicted for the other spatial resolutions.

Fig. 3 shows differently weighted modularities of the clustered networks as introduced in Section 2.3. While the non-weighted modularity of the *k-medoids Dijkstra* clustered network is significantly higher at all resolutions (9% in a 300-bus system) and increases at higher spatial resolutions, the difference for the electrically weighted modularity is rather small. In this case, although the *k-medoids Dijkstra Clustering* leads to higher modularities, both clustering approaches produce networks with very high modularities close to 1. This high order of magnitude is produced by a relatively high number of short intra-cluster lines with low reactances. In the context of the maximum resolution in this work of 700 buses and the resulting minimum cluster-to-cluster distances of 2.5 km (*k-medoids Dijkstra Clustering*; 5.6 km with *k-means Clustering*), these short lines, within or near substations, are hardly relevant, but receive high weights. To reduce this distortion, we also calculated the modularity excluding the quartile of the shortest lines (i.e. lines shorter than 350 m). As a result, the absolute values decrease, while the difference becomes clearer and generally increases at higher spatial resolutions (up to 12% in a 700-bus system).

Furthermore focusing on the 300-bus system, the generally stated lower degree of inter-cluster meshing can also be discovered when analysing Fig. 4. This holds true for the inner-German grid structure, whereas there are no differences between the clustered networks regarding the connections to foreign countries as these are modelled in a significantly lower spatial resolution. To deepen the analysis of the inner-German grid structure, the interconnections across the former inner-German border between the former FRG and former GDR are highlighted for the original network as well as for the clustered networks in Fig. 4. In the original network, due to the historic growth of the electrical network, in total, there are eight lines crossing the former border. The total capacity of those lines is about 19.2 GVA. The *k-medoids Dijkstra Clustering* leads to six aggregated crossing lines with a total capacity of about 19.5 GVA. In contrast, the *k-means Clustering* produces a clustered network with 22 aggregated lines crossing the former border, adding up to a capacity of about 40.4 GVA. This implies a significant overestimation of capacity of about 110%, whereas the novel approach estimates the original total crossing capacity more accurately.

Further investigations of those lines crossing the former border reveal the special importance of the HV level in this context. Notwithstanding the significant overestimation at the eHV level, the stub lines and stub meshes within the HV level in particular contribute to the overestimation of the total transmission capacity, as shown in Table 2. In the original network, about 7% of the corresponding transmission capacity is located in the HV level. Whereas the *k-medoids Dijkstra Clustering* represents this ratio with a value of 8% quite accurately, in the *k-means* clustered network, 33% of the total transmission capacity across the former border are located in the HV level.

The overestimation of crossing capacity in the case of the *k-means Clustering* remains for lower and higher spatial resolutions, as it can be seen in Table 2. For low spatial resolutions, overestimation can be observed in both clustering approaches, whereas for higher resolutions, the errors tend to decrease. When reducing to 100 buses or higher spatial resolutions, the overestimation using a *k-medoids Dijkstra Clustering* is negligibly small (≤1%) whereas the error using a *k-means Clustering*

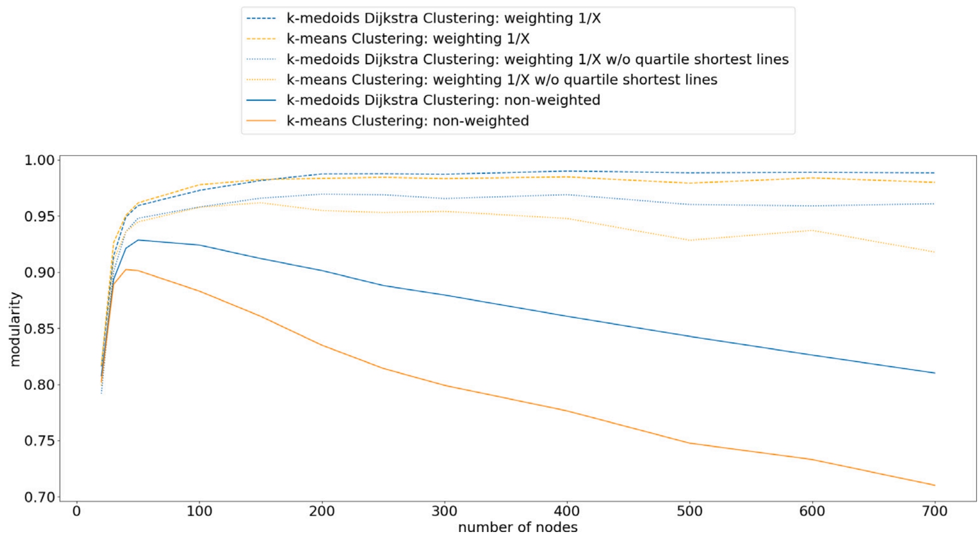
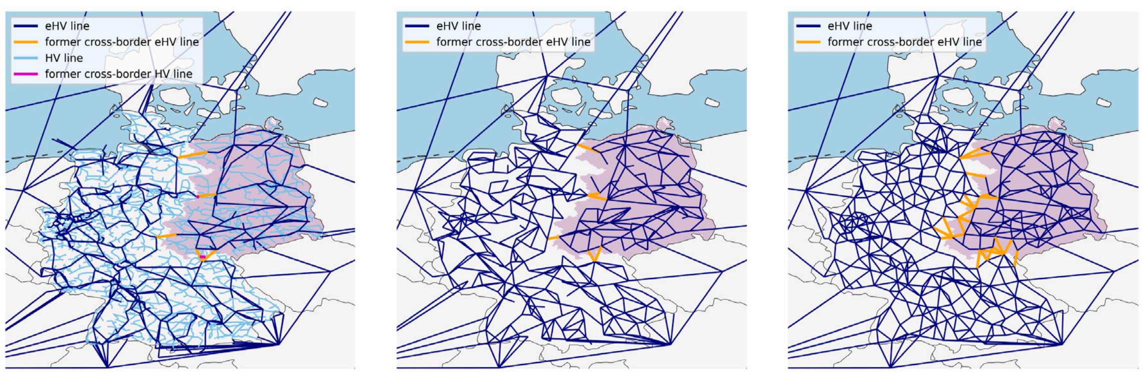


Fig. 3. Modularity of clustered networks for different spatial resolutions considering either no specific weighting of lines (as Clauset et al. [29]) or an electrical weighting (as Biener and Garcia Rosas [15]) i.e. the reciprocal value of the reactance. In addition, the electrically weighted modularity excluding the quartile of the shortest lines is shown.



(a) (b) (c)

Fig. 4. Network topologies of the original network including the eHV and HV grid (a) and a 300-bus network clustered by the *k-medoids Dijkstra Clustering* (b) as well as clustered by the *k-means Clustering* (c). The lines in the clustered networks are abstracted aggregated lines converted to eHV level. Lines crossing the former inner-German border between former FRG and GDR are highlighted.

is significantly higher in all spatial resolutions and does not go below 33%.

3.2. Optimisation results

Although there are only minor differences in the overall objective (i.e. annual system costs), network expansion needs, as well as the distribution of network and storage expansion differ significantly in the *k-medoids Dijkstra* from the *k-means* clustered systems.

As it can be seen in Table 3 under 'annual optimisation', which summarises the optimisation results for the 300-bus systems, the annual system costs differ negligibly showing about 0.5% higher costs in the *k-medoids Dijkstra* clustered network. The difference in the composition between these, i.e. overall operational and investment costs, is also negligible. The largest share of investment costs is battery expansion costs which are slightly lower in the *k-medoids Dijkstra* clustered network, while grid expansion costs are only a marginal part of the total investment costs. However, a significant relative difference can be observed when comparing the costs of grid expansion within Germany. In particular, the results implying a *k-medoids Dijkstra Clustering* show 16.7% higher costs.

In a worst-case scenario, if the system is optimised for only one snapshot, it is possible to generate results for the system in full spatial complexity. For this case, the optimised results for the different underlying clustering approaches can be compared to this benchmark (cf. Table 3, 'worst-case optimisation'). In line with the results of the comparative study of the annual optimisation results, significant differences between the two approaches only occur with respect to the costs of grid expansion within Germany. Whereas a *k-means Clustering* leads to a significant underestimation, the *k-medoids Dijkstra Clustering* leads to a slight overestimation.

These observations generally hold true when analysing the results for other spatial resolutions. In Fig. 5, the annual system costs, including marginal and investment costs, are plotted as bars for different spatial resolutions. The bars show that the costs for grid expansion makes up only a small part of the total cost. Furthermore, lines are plotted showing the relative differences between the systems reduced by the *k-medoids Dijkstra Clustering* and the *k-means Clustering* in several of the cost shares that add up to the total cost. As previously described, there are only minor differences in the individual shares, with the exception of the costs for inner-German grid expansion. The relative difference between *k-medoids Dijkstra* and the *k-means* clustered system is quite high for low spatial resolutions and is subject to large

Table 3
Comparison of the optimisation results for the underlying clustering approaches considering 300 buses. On the left side, annual optimisation results derived from the *k-medoids Dijkstra Clustering* approach are compared to the ones generated by the *k-means Clustering* method. The relative difference is calculated according to Eq. (5). On the right side, worst-case optimisation results for one hour of the year characterised by the highest overall absolute line loading (derived from the annual optimisation) considering both clustering approaches are compared to the results of the optimised network in full spatial complexity.

Cost type	Annual optimisation			Worst-case optimisation	
	Annualised costs in 10^9 €/a		rel. difference	rel. error to full complex model	
	k-means Clustering	*k-medoids Dijkstra Clustering*		*k-means Clustering*	*k-medoids Dijkstra Clustering*
Overall system	25.70	25.82	+ 0.5%	− 1.1%	+ 0.04%
Operational	16.36	16.68	+ 2.0%	− 0.9%	+ 0.2%
Expansion	9.35	9.15	− 2.1%	− 8.0%	− 5.3%
- inner-German grid	0.06	0.07	+ 16.7%	− 33.1%	+ 4.4%
- foreign grid	0.999	1.00	+ 0.1%	− 19.1%	− 18.1%
- battery	8.28	8.07	− 2.5%	–	–

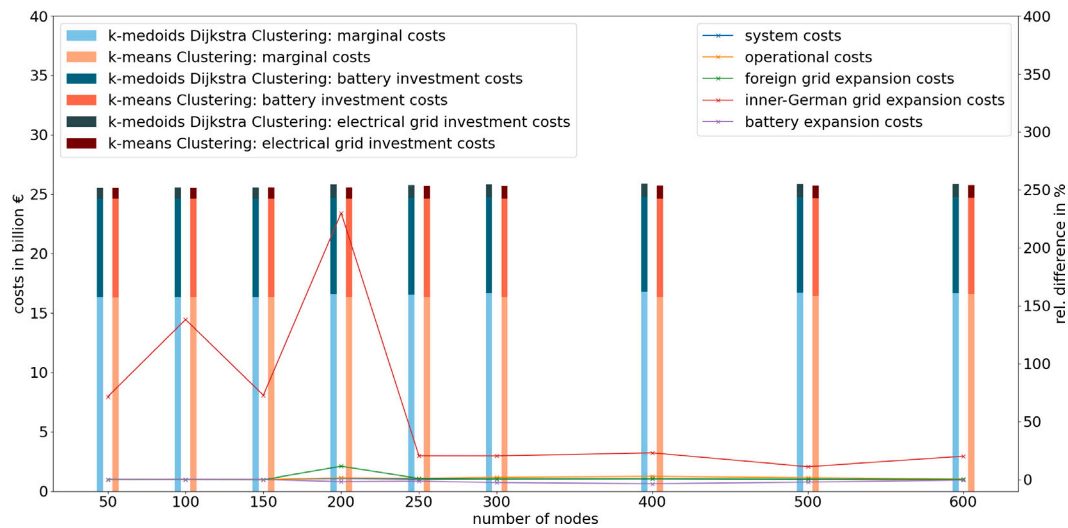


Fig. 5. Annual system costs and their composition for the underlying clustering approaches in different spatial resolutions. The relative difference is calculated according to Eq. (5).

fluctuations, but converges to about 15% as the number of considered buses increases.

In order to analyse how the results change for systems with different network expansion needs, several sensitivity variants with different specific capital cost assumptions have been calculated. As in the previous results, only inner-German grid expansion needs differ significantly between the *k-medoids Dijkstra* and the *k-means* clustered systems. Varying specific capital costs of lines results in higher relative differences (up to 42%) for systems with higher expansion needs and lower relative differences (down to 9%) for systems with lower expansion needs (cf. Fig. 6). Furthermore, Fig. 6 shows that grid expansion needs and the relative differences exhibit a reduced sensitivity to changing specific capital costs of batteries. However, an increase in the specific capital costs of batteries results in lower network expansion needs and a slight increase in relative differences between *k-medoids Dijkstra* and the *k-means* clustered systems.

Another sensitivity analysis has shown that the optimisation results react sensitively to neglecting the HV level. By utilising the *ehv Clustering* (cf. Section Data availability), all HV components are aggregated to their nearest eHV bus by using a *Dijkstra's algorithm*. Subsequently, the two distinct clustering approaches are used as described in Section 2 to explicitly cluster the eHV grid down to 300 buses. Consequently, the inner-German grid expansion in the *k-medoids Dijkstra* and *k-means* clustered systems differs only marginally (by −3%, cf. Eq. (5)).

Focusing again on the base base setting, in the *k-medoids Dijkstra* clustered network, 74 out of 533 transmission lines show expansion needs, whereas in the *k-means* clustered network, only 56 of 669 lines are extended within the optimisation. As shown in Fig. 7 and quantified in Table 4, significantly higher expansion needs can be identified in the systems implying a *k-medoids Dijkstra Clustering* within the former GDR and in the lines crossing the former inner-German border. Considering the latter, there is almost no need for expansion in the *k-means* clustered system, whereas in the *k-medoids Dijkstra* clustered network, five out of six lines show expansion needs. However, the total optimised capacity of these lines crossing the former inner-German border is higher in the *k-means* clustered network.

In contrast, higher storage expansion needs can be observed within the former GDR in the system implying a *k-means Clustering*. As shown in Fig. 7, huge expansion needs are located in the far east, close to the borders with Poland and the Czech Republic. In the former GDR, the optimised capacity of batteries in the *k-means* clustered network is higher, while in the rest of Germany, the optimised capacity is higher in the *k-medoids Dijkstra* clustered network.

3.3. Calculation time

Since the aim of reducing the complexity is to reduce the time required for optimisation, the calculation time is analysed in the following. Fig. 8 shows the time required to optimise the systems as a function of their spatial resolution. While the execution of the algorithms to reduce the spatial complexity must also be considered, it only accounts for a small fraction of the total computation time. For the highest considered spatial resolution, it takes about 45 min to conduct the *k-means Clustering* which is the same for the *k-medoids Clustering*. The time required to perform the *Dijkstra's algorithm* when clustering to 600 nodes is about 15 min. The sum of these computation times is less

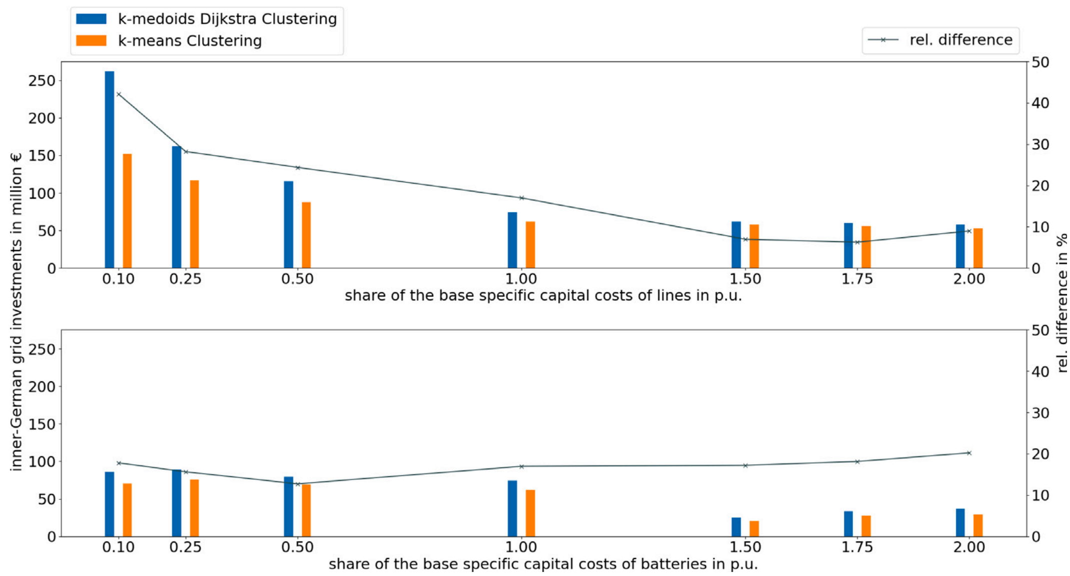


Fig. 6. Inner-German grid expansion needs for sensitivity variants with different assumptions on specific capital costs of lines and batteries in the system with 300 buses. In the upper panel, the resulting total capital costs of lines have been adjusted by the capital cost assumptions from the base setting in order to obtain comparable results and to account for cost differences at different voltage levels. The relative difference is calculated according to Eq. (5).

Table 4
Overview over the distribution of lines with expansion needs as well as expanded batteries in Germany for the underlying clustering approaches considering 300 buses.

	k-medoids Dijkstra Clustering		*k-means Clustering*	
Network expansion needs	Network expansion in TVA· km	Optimised capacity in TVA· km	Network expansion in TVA· km	Optimised capacity in TVA· km
in former FRG	3.3	37.6	3.4	46.4
in former GDR	1.1	13.6	0.2	14.3
crossing former border	0.4	1.3	0.005	2.05
Battery expansion needs	Expansion in GW	Optimised power in GW	Expansion in GW	Optimised power in GW
In former FRG	68.9	82.9	58.6	72.6
In former GDR	12.4	15.2	26.2	29.0

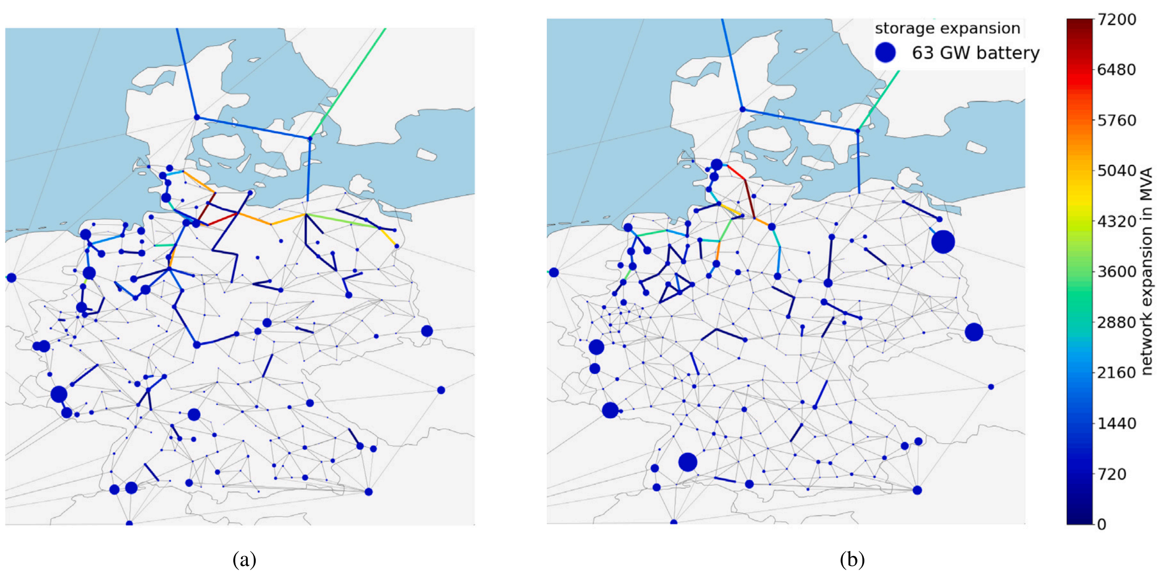


Fig. 7. Spatial distribution of network and storage expansion in 300-bus systems clustered by the *k-medoids Dijkstra Clustering* (a) as well as (b) by the *k-means Clustering*. The legends on the right apply to both figures.

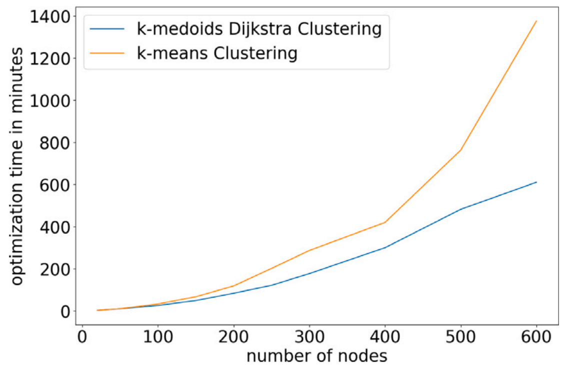


Fig. 8. Optimisation time for different spatial resolutions.

than 10% of the time required for optimisation when using a *k-medoids Dijkstra Clustering*. Therefore, this is not examined any further.

For the considered scale, the optimisation time increases with increasing spatial resolution for both, *k-medoids Dijkstra* and *k-means* clustered networks. At low spatial resolutions, the optimisation time using a *k-medoids Dijkstra Clustering* or a *k-means Clustering* does not show much differences. Instead, at higher resolutions, the time required to optimise the *k-medoids Dijkstra* clustered system is significantly less than the time required to optimise the *k-means* clustered system. For 600 nodes, the optimisation time in the *k-means* clustered network is more than twice as long.

4. Discussion

4.1. Discussion of results

The work presented provides strong and manifold evidence that the *k-medoids Dijkstra Clustering* avoids false inter-zonal meshing during the clustering process and is therefore able to represent the original network more accurately than the *k-means Clustering*. The degree of meshing and the transmission capacities are more accurately represented, which becomes particularly clear when looking at the former border between former FRG and GDR. Here, in particular at the HV level, the false inter-zonal meshing leads to a significant overestimation of transmission capacity when using a *k-means Clustering*. Similar findings can be derived from the modularity analysis. Fig. 3 shows a characteristic curve with similar values (considering the chosen spatial resolutions) as determined in other works (cf. [15,17]). The non-weighted as well as electrically weighted modularity values imply that the original network is more accurately represented when using a *k-medoids Dijkstra Clustering*. Similar to the findings of Biener and Garcia Rosas [15] on the relation between modularity and optimisation error, this implies a higher accuracy in the optimisation results when using networks clustered by the *k-medoids Dijkstra Clustering*.

The analysis of the optimisation results reveals only marginal differences in the total system costs. Only inner-German network expansion costs, which are higher in the *k-medoids Dijkstra* clustered network, differ significantly. This appears to be in line with the findings of the analysis of the clustering results, where a significant overestimation of transmission capacities can be observed in the *k-means* clustered networks. Consequently, the need for network expansion is underestimated in the optimisation. A similar observation regarding the detection of congested lines is made in [15], where the number of congested lines is underestimated in the *k-means* clustered networks.

The relative differences in inner-German network expansion costs are very high and vary significantly at low spatial resolutions, but tend to converge to a moderate value of around 15% in the higher and in our modelling more relevant spatial resolutions. Regarding the higher differences at lower spatial resolutions, Biener and Garcia Rosas [15] show an opposite tendency where fewer differences are observed when reducing to lower numbers of buses. However, in [15], both the number of buses in the original network as well as the degree of reduction are much lower than in this work, since in their analysis of congested lines, the reduced networks still contain at least 20% of the original number of buses. Furthermore, in our work, the costs of network expansion within Germany are quite low at low spatial resolutions, which makes the relative difference more sensitive to volatility. Consequently, in this work we cannot derive robust results for low spatial resolutions regarding differences in grid expansion needs.

In contrast, when analysing the optimisation results for systems of higher spatial resolutions, i.e. for our standard 300-bus system, a sensitivity analysis has indicated the robustness of our findings. In particular, four general observations can be emphasised. First, our findings particularly apply for systems including sub-transmission grid structures. Second, for systems with similar grid expansion needs, the relative differences between the two considered approaches (*k-medoids Dijkstra Clustering* and *k-means Clustering*) remain relatively constant. Third, in scenario variants with low grid expansion needs, energy has to be generated and used rather locally. Therefore, in particular the HV grid and its expansion become more relevant. Here, the relative differences slightly increase up to 20%. Fourth, when both, transmission and sub-transmission grid expansion become highly relevant, the *k-medoids Dijkstra* clustered systems show up to 42% higher grid expansion needs. In the light of ambitious German grid development plans (such as German Transmission System Operators [30]) and the general state of the art of the German energy transition, these significant differences are particularly important.

In the following, specific drivers that contribute to these significant overall differences in inner-German network expansion needs are highlighted. The underestimation of network expansion needs (when using the *k-means Clustering*) is particularly dominant for lines crossing the former border between former FRG and GDR as well as within the area of the former GDR. This, again, is related to the substantial overestimation of transmission capacity across the former border of 110% in the *k-means* clustered network. In contrast, the *k-medoids Dijkstra Clustering* leads to very accurate border crossing line capacities, leading to substantial grid expansion needs of 7 GVA, whereas the *k-means* clustered network leads to hardly any of such needs. The transmission capacity of the not yet optimised, *k-means* clustered network is already substantially higher than the optimised capacity in the *k-medoids Dijkstra* clustered network. In other words, the false inter-zonal meshing effect of about 21 GVA outnumbers by far the economically feasible expansion needs once the network is clustered more accurately. As a result, the eastern and western systems operate more isolated from each other, resulting in a substantially different distribution of network and storage expansion needs. In particular, this leads to higher network expansion requirements within the former GDR, while storage expansion needs are substantially lower. The reason for the relatively high storage capacities in eastern Germany within the *k-means* clustered system has been linked to loop flow behaviour, which forces energy to flow from north-western Europe (especially Denmark, where there is a high level of wind generation) towards eastern Germany and then back to southern load centres in Germany via eastern neighbouring countries (cf. [14]). Once the eastern parts are modelled electrically more isolated, due to a different topological distribution of the reactances, the power flow is less likely to flow via this described eastern route. Consequently, investments in storage units in eastern Germany are less profitable. To a moderate extent, the need for storage expansion is shifted to the rest of Germany although the differences are less substantial here as the generation and load centres are more directly and efficiently connected.

Addressing the issue of false inter-zonal meshing is particularly important because of the historically grown topology of the German power grid, which includes many stub lines and stub meshes (especially on HV level) along the former inner-German border. In general, wherever electrical obstacles or bottlenecks are part of the networks, e.g. caused by rivers, mountains or former or current political borders, a significant overestimation of transmission capacities and therefore a substantial underestimation of network expansion needs can be expected when applying a clustering method that hardly considers the original topological relations between buses and lines, such as the *k-means Clustering*. This is true for a wide range of highly relevant spatial resolutions. As a result, the spatial distribution of network and storage expansion needs can be predicted more accurately if important topological network relations are taken into account, as in the *k-medoids Dijkstra Clustering*, since structural differences may occur in the reduced networks, leading to misrepresented power flows.

The higher degree of accuracy does not come with the drawback of slower calculation times. Instead, the *k-medoids Dijkstra Clustering* outperforms the *k-means Clustering* in terms of total calculation times, although the clustering itself takes longer. This is due to the fact that, especially at high spatial resolutions, the time required to optimise the clustered systems, which accounts for the majority of the total calculation time, is significantly lower when applying a *k-medoids Dijkstra Clustering*. Most likely, the smaller number of lines and the lower degree of meshing lead to a smaller number of optimisation variables, reducing the solver's time to find the optimal solution. More complex modelling, e.g. including sector-coupling as in [18], is more likely to be successful due to reduced optimisation time and increased accuracy.

4.2. Limitations

The presented method considers the original network topology by taking into account the lengths of the lines, whereas other approaches use electrical parameters like reactance and admittance instead (as described in Section 1). Although line reactance and admittance are more important for modelling power flows, the use of the line lengths can be advantageous. Firstly, the use of line lengths limits the dependence on assumptions of electrical parameters. This is particularly beneficial in the area of open data modelling, as topologies can be modelled more accurately than the electrical parameters of each specific line. Furthermore, during optimisation, the reactance and admittance of lines may change as they are extended. Finally, the reactance and admittance of a line are strongly influenced by its length, which is taken into account in our method. However, the voltage level, which is also an important parameter influencing assumptions on electrical parameters, is not considered in our approach. Therefore, lines in lower voltage levels, which would potentially have higher impedances, are not weighted accordingly. Although we are unable to determine the impact of this on the modelling results, the aim of this work is to avoid false inter-zonal meshing, which is successfully achieved.

The presented method can be applied to any other large-scale power grid model. In general, the usage of this approach is preferable to the application of the commonly used *k-means Clustering*. Especially where electrical obstacles or bottlenecks are part of the networks, the avoidance of false inter-zonal meshing is particularly important to correctly predict the spatial allocation of network and storage expansion needs. In addition, the relevance of the improvements achieved by applying the novel approach remains particularly significant for grid models with a certain level of detail. When considering networks with comparatively low initial spatial resolutions (e.g. European system with one node per country), the resulting changes are not relevant as the representation of the topology is not as important as in cases like the one presented. This has been particularly shown by aggregating the 110 kV level to the transmission grid level before applying the compared clustering methods. The difference between *k-medoids Dijkstra* and *k-means* clustered network is negligible in this case, since many stub lines and stub meshes occur in the 110 kV level and the reduction degree is much lower.

Due to the high spatial resolution of the model, full year optimisations with full spatial complexity were not performed as this is not possible with the given computational resources at a reasonable time. Instead, the method was compared with the commonly used *k-means Clustering*, as this method has been applied in many works before (cf. Section 1). Therefore, the complexity was reduced to different spatial resolutions between 50 and 600 buses. 600 buses only account for about 6.5% of the original number of buses, and therefore calculations are only performed within a relatively narrow resolution band compared to the number of buses in the full complex model. However, since the original model is so detailed, it already results in a high spatial resolution compared to other works dealing with the German electricity grid (e.g. [15]). Moreover, converging comparative results at higher spatial resolutions, which show significant differences, argue for an adequate model setup. Finally, the underlying clustering results are compared to the original network topology. The results can be linked to our optimisation results as well as discussed in the light of recent studies.

A worst-case scenario was developed to allow a comparative study on the optimisation results for the original system at full spatial complexity. The examination of extreme cases is a common approach in power grid planning. Although the general model setup and tool are essentially not designed for such static worst-case calculations, the comparison increases the reliability of the results presented in this paper. Likewise, the performed sensitivity analyses demonstrate the robustness of the investigation showing substantial differences in inner-German grid expansion needs when applying the considered methods.

Although we expect a significant impact of our novel method when modelling a sector-coupled energy system, as already done in [18], we neglected sector coupling in this work. Consequently, the effects in an even more complex sector-coupled model cannot be explicitly discussed.

In this work, a scenario for the year 2035 is used. For the optimisation of a scenario including a 100% RE generation, even higher network and storage expansion needs are expected (cf. [18]). Therefore, it is even more important to avoid false inter-zonal meshing in order to adequately predict expansion needs and their allocation. We therefore expect that the presented method will become increasingly important when modelling 100% RE systems.

Future work should address the limitations of this work. Of particular interest will be the consideration of electrical distances, the performance of new calculations with higher or full spatial complexity (e.g. by further reducing the temporal dimension or focusing on a smaller regional scope), and comparative analyses in the context of a sector-coupled model and a 100% RE scenario.

5. Conclusion

In this paper, a *k-means Clustering* is combined with a *Dijkstra's algorithm* to take into account actual path lengths between the original buses and their representatives when clustering a power grid model. The novel method is applied to a spatially detailed model of the German transmission and sub-transmission grid to avoid false inter-zonal meshing. The effects on the clustered grids and the corresponding optimisation results are analysed through a comparative study referring to the commonly used *k-means Clustering*.

The analysis reveals that the network topology and transmission capacities are more accurately modelled by the novel approach. Avoiding false inter-zonal meshing is particularly important where geographical or political borders lead to many stub lines and stub meshes (especially on HV level). Otherwise, when using a *k-means Clustering*, a substantial overestimation of meshing and transmission capacity results in structural changes in the clustered networks, a moderate to high underestimation of the overall German network expansion needs, and a different spatial allocation of network and storage expansion needs. In addition to a moderate to high impact on the aggregated national level (20% less inter-cluster lines and up to 41% higher grid expansion needs in a 300-bus system), substantial differences can be observed on a regional level. In particular, the inter-zonal lines crossing the former inner-German border are modelled very accurately in the novel approach, whereas in the *k-means Clustering*, a substantial overestimation of 21 GVA outnumber the expansion needs when applying the *k-medoids Dijkstra Clustering* by far.

Hence, although there are no significant differences in the annual system costs, as network expansion costs are only a small part of these, it is recommended to use a clustering method that takes into account the original topological relations between buses and lines in order to model the grid topology and its regionally distributed expansion needs accurately. Finally, the positive side-effect of reduced optimisation times opens up new possibilities in terms of the complexity of calculations in other dimensions, e.g. calculations including more technologies, as required in highly relevant sector-coupled models.

CRediT authorship contribution statement

Katharina Esterl: Writing – original draft, Writing – review & editing, Conceptualization, Visualization, Software. **Carlos Andrés Epia Realpe:** Writing – review & editing, Visualization, Software. **Ulf Philipp Müller:** Writing – review & editing, Conceptualization.

Funding

The authors thank the Federal Ministry for Economic Affairs and Climate Action for funding the research project *eGo^n* (grant number: 03EI1002).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Objective function of LOPF

Within the LOPF in *Python for Power System Analysis* (*PyPSA*), the objective function includes the minimisation of the annual system costs defined as the sum of annualised network and storage expansion costs and annual dispatch costs (cf. Eq. (3)). A detailed mathematical description is given in [6].

$$\min_{F_\ell, H_{n,s}, g_{n,r,t}, h_{n,s,t}} \Bigg[\underbrace{\sum_{\ell} c_\ell F_\ell}_{\text{grid expansion costs}} + \underbrace{\sum_{n,s} c_{n,s} H_{n,s}}_{\text{storage expansion costs}} + \underbrace{\sum_{n,r,t} (w_t \cdot o_{n,r} \cdot g_{n,r,t})}_{\text{generator dispatch costs}} + \underbrace{\sum_{n,s,t} (w_t \cdot o_{n,s} \cdot [h_{n,s,t}]^+)}_{\text{storage dispatch costs}} \Bigg] \quad (3)$$

ℓ: index passive branch (AC-line)
n: index bus
r: index generator
s: index storage
t: snapshot
c_ℓ: CAPEX passive branch
F_ℓ: capacity passive branch
w_t: snapshot weighting
$o_{n,r}$: OPEX of generator *n, r*
$g_{n,r,t}$: dispatch of generator *n, r, t*
$c_{n,s}$: CAPEX of storage *n, s*
$H_{n,s}$: capacity of storage *n, s*
$o_{n,s}$: OPEX of storage *n, s*
$h_{n,s,t}$: dispatch of storage *n, s, t*

Appendix B. Modularity

The modularity Q, as originally developed in [29], is used to examine how well the clustered networks depict the meshing in the original network and is defined in Eq. (4) [17]:

$$Q = \frac{1}{2m} \sum_{v,w} (A_{vw} - \frac{k_v k_w}{2m}) \delta(c_v, c_w), \quad (4)$$

where v and w are buses, A_{vw} is the weighted adjacency matrix of the network graph, G, m the sum of all edge weights in the graph and k_v the weighted degree of bus v. The quantities are defined as

$$A_{vw} := \begin{cases} w_{(v,w)} & \text{if } (v,w) \in E \\ 0 & \text{otherwise} \end{cases}, \quad m := \frac{1}{2} \sum_{v,w} A_{vw}, \quad k_v := \sum_{w} A_{vw}$$

The delta function is given as described below. c_v denotes the cluster bus v which it is assigned to.

$$\delta_{(c_v, c_w)} := \begin{cases} 1 & \text{if } c_v = c_w \\ 0 & \textit{otherwise} \end{cases}$$

Clauset et al. [29] originally introduce the modularity without weighting of the connections between nodes. In the work by Biener and Garcia Rosas [15] and Frysztacki et al. [17], a measure of electrical distance is chosen for the weighting. While Biener and Garcia Rosas [15] use the reciprocal value of the reactance (i.e. $|\frac{1}{x_{v,w}}|$), Frysztacki et al. [17] choose the admittance $|y_{v,w}|$ of each line (v, w). It should be noted that, due to the focus on LOPF calculations, where the linearisation of the power flow results in the reactance X being the relevant parameter, we utilise the weighting as defined by Biener and Garcia Rosas [15].

Appendix C. Definition of relative difference

For the comparative analysis, the relative difference is defined as follows:

$$\delta r_a = \frac{r_a^{Dijkstra} - r_a^{kmeans}}{r_a^{Dijkstra}} \cdot 100\% \quad (5)$$

δr_a: relative difference of attribute a in %
$r_a^{Dijkstra}$: result of attribute a in a *k-medoids Dijkstra* clustered network
r_a^{kmeans}: result of attribute a in a *k-means* clustered network

Appendix D. Acronyms

eTraGo *electrical Transmission Grid Optimization*

PyPSA *Python for Power System Analysis*

eHV extra-high voltage

FRG Federal Republic of Germany

GDR German Democratic Republic

HV high voltage

LOPF Linear Optimal Power Flow

OEP *Open Energy Platform*

OSM OpenStreetMap

RE renewable energies

TYNDP Ten-year network development plan

Data availability

This work has been developed as part of the research project *eGo^n* following open source and open data principles.

•The data model will be available via the *Open Energy Platform* (*OEP*): https://openenergy-platform.org/. It can be downloaded here: https://zenodo.org/records/8376714
•The tool *eTraGo* is published on *Github*. The latest release contains all the described functionalities including the *k-medoids Dijkstra Clustering*: https://github.com/openego/eTraGo/tree/0.9.0
•The installation process of *eTraGo*, its functionalities as well as how to use *eTraGo* are explained and examples for its application are given here: https://etrago.readthedocs.io/en/latest/index.html
•The parameterisation used for the calculations conducted within this work is summarised here: https://github.com/openego/eTraGo/blob/paper/kmedoids-dijkstra-2/etrago/args.json
•The implementation of the introduced method (*k-medoids Dijkstra Clustering*) can be found here: https://github.com/openego/eTraGo/blob/58b953fa3a2a51e30855d9ee73e9758b8699a2a6/etrago/cluster/spatial.py#L681
• The implementation of the function aggregating the HV level to the eHV level (*ehv Clustering*): https://github.com/openego/eTraGo/blob/58b953fa3a2a51e30855d9ee73e9758b8699a2a6/etrago/cluster/electrical.py#L426

References

[1] T. Brown, D. Schlachtberger, A. Kies, S. Schramm, M. Greiner, Synergies of sector coupling and transmission reinforcement in a cost-optimised, highly renewable European energy system, Energy 160 (2018) 720–739, http://dx.doi.org/10.1016/j.energy.2018.06.222, URL: https://www.sciencedirect.com/science/article/pii/S036054421831288X.

[2] H.C. Gils, H. Gardian, J. Schmugge, Interaction of hydrogen infrastructures with other sector coupling options towards a zero-emission energy system in Germany, Renew. Energy 180 (2021) 140–156, http://dx.doi.org/10.1016/j.renene.2021.08.016, URL: https://www.sciencedirect.com/science/article/pii/S0960148121011769.

[3] M. Hoffmann, L. Kotzur, D. Stolten, M. Robinius, A review on time series aggregation methods for energy system models, Energies 13 (3) (2020) http://dx.doi.org/10.3390/en13030641, URL: https://www.mdpi.com/1996-1073/13/3/641.

[4] V. Aryanpur, B. O'Gallachoir, H. Dai, W. Chen, J. Glynn, A review of spatial resolution and regionalisation in national-scale energy systems optimisation models, Energy Strategy Rev. 37 (2021) 100702, http://dx.doi.org/10.1016/j.esr.2021.100702, URL: https://www.sciencedirect.com/science/article/pii/S2211467X21000882.

[5] R. Dashti, M. Rouhandeh, Power distribution system planning framework (a comprehensive review), Energy Strategy Rev. 50 (2023) 101256, http://dx.doi.org/10.1016/j.esr.2023.101256, URL: https://www.sciencedirect.com/science/article/pii/S2211467X23002067.

[6] J. Hörsch, T. Brown, The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios, in: 2017 14th International Conference on the European Energy Market, EEM, 2017, http://dx.doi.org/10.1109/EEM.2017.7982024, URL: https://ieeexplore.ieee.org/document/7982024/.

[7] J. Hörsch, F. Hofmann, D. Schlachtberger, T. Brown, PyPSA-Eur: An open optimisation model of the European transmission system, Energy Strategy Rev. 22 (2018) 207–215, http://dx.doi.org/10.1016/j.esr.2018.08.012, URL: https://www.sciencedirect.com/science/article/pii/S2211467X18300804.

[8] L. Wienholt, U.P. Müller, J. Bartels, Optimal sizing and spatial allocation of storage units in a high-resolution power system model, Energies 11 (12) (2018) http://dx.doi.org/10.3390/en11123365, URL: https://www.mdpi.com/1996-1073/11/12/3365.

[9] U. Müller, L. Wienholt, I. Cussmann, The role of the high voltage power level in future power systems and their modelling, in: 2018 International Conference on Smart Energy Systems and Technologies, SEST, 2018, pp. 1–6, http://dx.doi.org/10.1109/SEST.2018.8495721.

[10] U.P. Müller, B. Schachler, M. Scharf, W.-D. Bunke, S. Günther, J. Bartels, G. Pleßmann, Integrated techno-economic power system planning of transmission and distribution grids, Energies 12 (11) (2019) http://dx.doi.org/10.3390/en12112091, URL: https://www.mdpi.com/1996-1073/12/11/2091.

[11] M.M. Frysztacki, J. Hörsch, V. Hagenmeyer, T. Brown, The strong effect of network resolution on electricity system models with high shares of wind and solar, Appl. Energy 291 (2021) 116726, http://dx.doi.org/10.1016/j.apenergy.2021.116726, URL: https://www.sciencedirect.com/science/article/pii/S0306261921002439.

[12] F. Neumann, E. Zeyen, M. Victoria, T. Brown, The potential role of a hydrogen network in Europe, Joule 7 (8) (2023) 1793–1817, http://dx.doi.org/10.1016/j.joule.2023.06.016, URL: https://www.sciencedirect.com/science/article/pii/S2542435123002660.

[13] C. Reinert, B. Nilges, N. Baumgärtner, A. Bardow, This is SpArta: Rigorous optimization of regionally resolved energy systems by spatial aggregation and decomposition, 2023, http://dx.doi.org/10.48550/arXiv.2302.05222.

[14] U.P. Müller, Combined Optimization of Grid and Storage Expansion in the German Power System (Ph.D. thesis), Europa-Universität Flensburg, 2021, URL: https://d-nb.info/125765599X/34.

[15] W. Biener, K.R. Garcia Rosas, Grid reduction for energy system analysis, Electr. Power Syst. Res. 185 (2020) 106349, http://dx.doi.org/10.1016/j.epsr.2020.106349, URL: https://www.sciencedirect.com/science/article/pii/S0378779620301553.

[16] M. Weber, H.K. Çakmak, U. Kühnapfel, V. Hagenmeyer, A novel method for topology preserving static network reduction, in: 2023 IEEE 11th International Conference on Smart Energy Grid Engineering, SEGE, 2023, pp. 58–63, http://dx.doi.org/10.1109/SEGE59172.2023.10274576.

[17] M.M. Frysztacki, G. Recht, T. Brown, A comparison of clustering methods for the spatial reduction of renewable electricity optimisation models of Europe, Energy Inform. 5 (2022) http://dx.doi.org/10.1186/s42162-022-00187-7.

[18] C. Büttner, K. Esterl, I. Cußmann, C.A. Epia Realpe, J. Amme, A. Nadal, Influence of flexibility options on the german transmission grid — A sector-coupled mid-term scenario, Renew. Sustain. Energy Transition 5 (2024) 100082, http://dx.doi.org/10.1016/j.rset.2024.100082, URL: https://www.sciencedirect.com/science/article/pii/S2667095X24000060.

[19] German Transmission System Operators, Netzentwicklungsplan strom 2035, version 2021, 1. Entwurf, 2021, URL: https://www.netzentwicklungsplan.de/archiv/netzentwicklungsplan-2035-2021.

[20] Geofabrik GmbH and OpenStreetMap-Mitwirkende, OpenStreetMap data extracts, stand 01.01.2021, 2021, URL: https://download.geofabrik.de/.

[21] M. Scharf, Entwicklung Eines Modells des Deutschen Übertragungsnetzes auf Basis der Offenen Geodatenbank Openstreetmap (Bachelor's thesis), Fachhochschule Flensburg, 2015.

[22] W. Medjroubi, U.P. Müller, M. Scharf, C. Matke, D. Kleinhans, Open data in power grid modelling: new approaches towards transparent grid models, Energy Rep. 3 (2017) 14–21.

[23] U. Müller, L. Wienholt, D. Kleinhans, I. Cußmann, W.-D. Bunke, G. Pleßmann, J. Wendiggensen, The eGo grid model: An open source approach towards a model of german high and extra-high voltage power grids, J. Phys. Conf. Ser. 977 (2018) 012003, http://dx.doi.org/10.1088/1742-6596/977/1/012003.

[24] T. Brown, J. Hörsch, D. Schlachtberger, PyPSA: Python for power system analysis, J. Open Res. Softw. 6 (4) (2018) http://dx.doi.org/10.5334/jors.188, arXiv:1707.09913.

[25] X. Jin, J. Han, K-medoids clustering, in: C. Sammut, G.I. Webb (Eds.), Encyclopedia of Machine Learning, Springer US, Boston, MA, 2010, pp. 564–565, http://dx.doi.org/10.1007/978-0-387-30164-8_426.

[26] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms, third ed., The MIT Press, 2009.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (null) (2011) 2825–2830.

[28] A.A. Hagberg, D.A. Schult, P.J. Swart, Exploring network structure, dynamics, and function using networkx, in: G. Varoquaux, T. Vaught, J. Millman (Eds.), Proceedings of the 7th Python in Science Conference, Pasadena, CA USA, 2008, pp. 11–15.

[29] A. Clauset, M.E.J. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E 70 (2004) 066111, http://dx.doi.org/10.1103/PhysRevE.70.066111, URL: https://link.aps.org/doi/10.1103/PhysRevE.70.066111.

[30] German Transmission System Operators, Netzentwicklungsplan strom 2035, version 2023, 2. Entwurf, 2023.