House of Energy Markets & Finance

Identifying key elements for adequate simplifications of investment choices – the case of Wind energy expansion

HEMF Working Paper No. 01/2021

by

Arne Pöstges

and

Christoph Weber

January 2021

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Identifying key elements for adequate simplifications of investment choices – The case of wind energy expansion by Arne Pöstges and Christoph Weber

Abstract

The analysis of future energy systems with increasing shares of renewable energy production poses various challenges to models used in the field of energy system analysis. Aggregation is one solution to reduce the computation time of large optimization problems, especially for optimization models with endogenous capacity expansion. Since the economic viability of renewable investments is not directly driven by physical and technical characteristics but rather by the corresponding revenue and cost streams, we propose a novel aggregation method. The aggregation covers both the spatial and the technological dimension of wind investment choices, i.e. the siting as well as the turbine choice. Four value components are defined, related e.g. to the total yield or the site-specific infeed profile of investment choices. A clustering approach is applied to these value components to identify groups of investment choices that can be aggregated without excessive loss of accuracy. The approach is applied in a case study for the German electricity system and the influence of the different value components on the combined technological and spatial aggregation is analyzed.

Keywords: aggregation, clustering, value components, wind energy expansion

JEL-Classification: C43, C61, O21

Arne Pöstges
(*CORRESPONDING AUTHOR)
House of Energy Markets and Finance,
University of Duisburg-Essen, Germany
Universitätsstraße 2, 45141 Essen
+49-(0)201 / 183-6862
Arne.Poestges@uni-due.de
www.hemf.net

Christoph Weber
House of Energy Markets and Finance,
University of Duisburg-Essen, Germany
Universitätsstraße 2, 45141 Essen
+49-(0)201 / 183-2966
Christoph.Weber@uni-due.de
www.hemf.net

Content

1 Introduction

The increasing share of renewable energy production in future energy systems poses multiple challenges to energy modelers. Among other things, complexity and model size lead to increasing computation times, especially for optimization models with endogenous capacity expansion. Reducing the computation time by aggregation requires a comprehensive understanding of factors that affect the resulting optimal expansion plan. As optimizing energy system models are frequently used to inform policy makers or even help shaping policies, it is equally important to get an improved understanding of the factors that drive differences between model results and reality.

Multiple propositions have already been made to reduce model complexity or helping to solve large-scale models. Examples of the latter are decompositions methods like Benders decomposition (cf. Bloom, 1983) and Dantzig-Wolfe (cf. e.g. Sowa et al., 2016) decomposition. In the realm of model complexity reduction, timeseries aggregation has been particularly investigated, cf. the recent review by Hoffmann et al. (2020). Much less emphasis has been devoted to investigating the impact of spatial and technological aggregation.

One reason is certainly that spatial aggregation has a two-dimensional, inhomogeneous underlying as starting point, with contours of regions and countries frequently shaped by natural phenomena like rivers or shorelines. And for technologies, focusing on one representative technology has seemed adequate as long as mostly controllable power generation units like, coal, nuclear or gas units were built, where economies of scale in cost and heat rate were present yet at the same time plant size was limited by engineering constraints. With increasing focus on renewable investment both the location of the generation units and their technology characteristics become more important as they shape the infeed characteristics and thus the economic viability of investments. Technology characteristics are particularly diverse for wind energy, where hub height and the dimensioning of the electric generator relative to the rotor size have an impact on overall yield and profile of infeed (cf. e.g. Hirth and Müller, 2016, Bucksteeg, 2019 and Klie and Madlener, 2020).

At the same time the economic viability of renewable investments is not directly driven by physical and technical characteristics but rather by the corresponding revenue and cost streams. Therefore, the paper at hand proposes a novel method for aggregating both the spatial and the technological dimension of wind investment choices by considering value components of so-called investment choices. By investment choices we designate the combinations of sites and turbine types that are selected by investors, e.g. a so-called low-speed turbine in a mountainous region far from the coast or a low hub-height standard turbine on a coastal site.

We therefore develop subsequently an analytical approach to distinguish various value components, related e.g. to the total yield or the site-specific infeed profile of an investment choice. We then identify different model settings that allow us to assess numerically these value components. In turn, a clustering approach may be applied to these value components to identify groups of investment choices that can be aggregated without excessive loss of accuracy.

The remainder of the paper is structured as follows: section 2 presents an overview on the relevant literature, whereas section 3 is devoted to the developed methodology. The application is then presented in section 4 and conclusions are presented in section 5.

2 General considerations and relevant literature

The size and complexity of electricity market models constantly increases, mainly driven by a rising share of renewables, leading to the challenge of computational tractability.

Broadly speaking this challenge is driven by two factors: On the one hand problem complexity through e.g. coupling of sectors and on the other hand problem details such as the necessarily high resolution of renewable data leading to the challenge of an increasing problem size and long computing times (cf. Pfenninger, 2017). Possible solutions to model complexity are inter alia reformulation, parallelization or simplification (e.g. ignoring restrictions such as reserve provision). Large problem sizes due to the amount of input data are most commonly addressed by decomposition and aggregation. Thereby three aspects strongly influence the problem size on the level of input data: physical resp. technological detail as well as spatial and temporal resolution.

Since location and technology specifications are crucial for the infeed profiles of renewable energy sources (RES), the focus is subsequently on the reduction of two problem dimensions: spatial resolution and technological detail. Increasing shares of RES in energy systems redirect the focus away from power yield to the economic value and system friendliness of the renewable infeed. Hence, several studies recently analyzed the influence of wind turbine type and location on energy systems, respectively the changes in system cost and value of generation technologies. Hirth and Müller (2016) analyze the effect of system friendly wind turbines on the market value of wind power and find that system friendly wind turbines (taller towers, smaller specific power) yield a higher annual average capacity factor and smoother output, which increase the long-term market value of wind power. Furthermore, Bucksteeg (2019) focusses on the effects of wind power diversification to support generation adequacy and Klie and Madlener (2020) contrast the influence of spatial diversification and design of wind turbines. Latter conclude that the cost savings due to turbine diversification are less important compared to the effect of system friendly turbines, a finding which aligns with some of our results.

A broad range of aggregation methods have been developed over the last two decades, nevertheless comparative analysis of these various methods in the field of electricity system models are still rare. Inter alia the former structure of analyzed energy systems and ongoing developments and changes of the technological and political framework have induced a focus on methods of temporal aggregation. Measured by the number of relevant publications, apparently, time series aggregation is the best analyzed field of aggregation compared to spatial and technological aggregation (Kotzur et al., 2018; Teichgraeber and Brandt, 2019). Hoffmann et al. (2020) present a review on the broad field of time series aggregation methods for energy

system models by categorizing applied aggregation methods in 130 publications between 1999 and 2019.

Furthermore, the lack of comparable work regarding spatial and technological aggregation may be explained as follows. In terms of technology there was no need to intensively consider renewable energy sources for decades, because of their relatively small importance given the small size of single units and their low number. But due to the increasing share of renewables in electricity systems, the need for aggregation e.g. of wind turbines to wind farms gained attention.

Zhang et al. (2013) consider spatio-temporal correlations of wind power to develop a cluster methodology for wind turbines and Gómez-Muñoz and Porta-Gándara (2002) identify representative daily wind patterns based on a two-step clustering approach. Slootweg and Kling (2003) as well as Ali et al. (2013) focus on the influence of technical detail on aggregation.

The analysis of spatial aggregation in the field of optimization-based electricity system models considering renewable power production seems to be even rarer than analysis regarding technological aggregation. But for instance Frew et al. (2016) analyzes cost benefits of geographical aggregation as flexibility mechanism to integrate highly volatile renewables into the United States power grid.

In terms of methodology, clustering is the most common instrument for temporal and spatial aggregation within optimization problems by far (cf. e.g. Aghabozorgi et al., 2015; Gómez-Muñoz and Porta-Gándara, 2002; Pineda and Morales, 2018; Saxena et al., 2017; Zatti et al., 2019; Zhang et al., 2013). The algorithms which are used most frequently are k-means (cf. e.g. Adhau et al., 2014; Bahl et al., 2017; Fazlollahi et al., 2014; Green et al., 2014), k-medoids (cf. e.g. Domínguez-Muñoz et al., 2011; Schütz et al., 2017; Stadler et al., 2014) and hierarchical clustering (cf. e.g. Nahmmacher et al., 2016; Pfenninger, 2017). In the field of technological aggregation more individual and less clustering based approaches are considered. This may be due to the necessary consideration of both generation characteristics and cost parameters to fully incorporate the technology design for good aggregation. The idea to base a clustering on economic values has already been applied for spatial aggregation in the context of bidding zone delimitation. Locational marginal prices (LMPs) are thereby utilized to identify and aggregate similar price zones (cf. Breuer and Moser, 2014; Burstedde, 2012; Felling and Weber, 2018).

Consequently, the paper at hand proposes a novel method for aggregating both the spatial and the technological dimension of wind investment choices. These are defined as possible combinations of sites (locations) and turbine types (technologies, e.g. strong vs. low wind turbines). Thereby it is proposed to base the aggregation on a consideration of multiple value components. The value of an investment choice is therefore decomposed additively into: 1. the yield value component (driven by the site-specific full load hours/capacity factor), 2. the resource

related value component (driven by the general market value factor of wind), 3. the technology specific value component (driven by the selected turbine type), 4. the spatial heterogeneity value component (driven by the wind profile of the selected site) and 5. the grid component (considering the network load and resulting nodal price differences). By grounding aggregation on these factors, we aim to reduce errors of first (selection of wrong technologies) and second (non-selection of right technologies) kind within subsequent applications within electricity system optimization[1].

Summarizing, the contribution of this work is threefold: Firstly, it helps closing the gap in spatial and technological aggregation. Secondly, the approach is complementing the standard methods of aggregation based on time-series clustering with an approach also based on economic parameters. Thirdly, it supplements recent work in the analysis of geographical and technological diversification of wind turbines within electricity system optimization.

[1] The minimization of these errors is in line with solving a meta-problem as discussed in Pöstges and Weber (2019). Since this meta-problem is unsolvable, the approach at hand approximates the solution by using multiple solvable sub-problems.

3 Methodology

In section 3.1, the general idea and an overview of the algorithm to identify suitable aggregation of investment choices are presented. In the following section 3.2, the analytics and an illustrative example of the decomposition of the value of an investment choice are described and subsequently the scenario definitions and the clustering algorithm are introduced in section 3.3 respectively 3.4.

3.1 General approach

As mentioned above, the starting point of the presented approach is the net present value of an investment choice which is a key parameter in the decision-making, both for individual investors and within an electricity system model. It is defined as the net present value, i.e. the sum of discounted cash flows which include the contribution margins over the operation period minus the (upfront) investment cost (cf. Eq. (1)).

$$NPV_{a,i} = \sum_{t} e^{-rt} Y_{t,a,i} p_t - c_i K_{a,i} \tag{1}$$

Single indicators directly related to site or technology, e.g. full load hours or investment costs, are not sufficient to capture key drivers of decision making or to identify general principles of aggregation. By contrast, the defined value of an investment choice reflects multiple spatial and technological factors. The value is decomposed further into several value components in order to include value drivers that are potentially important, yet with an importance that varies depending on the actual decision context. This allows to analyze their individual choice-dependent contribution and to develop an aggregation that provides suitable results in different decision contexts. In order to obtain such an aggregation, we propose an algorithm that includes five steps as shown in Figure 1.

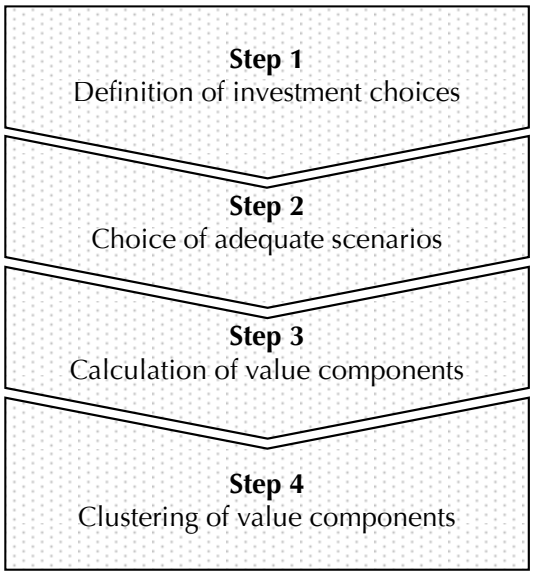

Figure 1: Overview of the algorithm

In a first step, the investment choices as objects for the clustering algorithm are defined. Therefore, spatial and technological aspects are considered simultaneously by defining the investment choices as combinations of site and (turbine) technology characteristics. Furthermore, distinct scenarios and hence relevant parameters such as prices and installed capacities are selected (cf. step 2, section 3.3) to compute the value components (cf. step 3). These value components are introduced in section 3.2 as they form the conceptual core of the methodology. The first three steps lay the foundations for the clustering carried out in step four. The clustering is split into the two sub-steps of predefining the starting point for a k-means clustering and computing the clusters itself (cf. section 3.4). This separation is useful because of the different suitability of cluster algorithms to define an appropriate number of clusters and to group the data.

3.2 Decomposition of the value of an investment choice

The value of an investment choice defined as the sum of all discounted cash flows (cf. Eq. (1)), may be decomposed additively into:

1. the **yield value component** which is driven by the site-specific full load hours,
2. the **resource related value component** which is driven by the general market value factor of the renewable technology,
3. the **technology specific value component**, driven by the selected turbine type,
4. the **site-specific value component**, driven by the infeed profile of the selected site and
5. for certain scenarios the **grid value component**, reflecting the network congestions and resulting nodal price differences (cf. Appendix A).

In section 3.2.1, the value components are derived analytically and an illustrative example is presented in section 3.2.2.

3.2.1 Analytics

Since it is assumed that investment choices may be evaluated based on one representative year (or possibly several alternative years), the value metrics shown in Eq. (1) is slightly modified. We focus on the annualized (net present) value AV based on the revenues of all timesteps t within one year, which allows us to drop the discounting factor e^{-rt}. The corresponding capital costs are obtained by multiplying the investment cost with the annuity factor a ($c_i^{ann} = c_i a$) in Eq. (2).

$$AV_{a,i} = \sum_t Y_{t,a,i} p_t - c_i^{ann} K_{a,i} \tag{2}$$

To enable the mathematical decomposition, an expanded representation of the production $Y_{t,a,i}$ is introduced in Eq. (3).

$$Y_{t,a,i} = \theta_{t,a,i} K_{a,i} \tag{3}$$

The production is replaced with the capacity $K_{a,i}$ times the i*ndividual capacity factor* $\theta_{t,a,i}$ which is defined as the fraction of installed capacity available for production depending on time, location and unit, multiplied with the capacity of technology type i at site a. The time dependent capacity factors represent the time series or infeed profiles of e.g. a wind turbine. In Table 1 the averaged[2] capacity factors utilized in the further steps of the decomposition (right hand part of eq. (10) and below) are defined mathematically and briefly explained.

Table 1: Definition and interpretation of capacity factors

Definition	Description
$\theta_{t,a,i}$ (4)	*Individual capacity factor* (per time step, site and technology)
$\bar{\theta}_{t,a,\cdot} = \frac{\sum_i (\theta_{t,a,i} * K_{a,i})}{\sum_i (K_{a,i})}$ (5)	*Average portfolio capacity factor* (per time step and site, averaged over all technologies)
$\bar{\theta}_{\cdot,a,\cdot} = \frac{\sum_t \bar{\theta}_{t,a,\cdot}}{N_T} = \frac{\sum_{t,i} (\theta_{t,a,i} * K_{a,i})}{N_T \sum_i (K_{a,i})}$ (6)	*Average time and portfolio capacity factor* (per site, averaged over time and all technologies)
$\bar{\theta}_{t,\cdot,\cdot} = \frac{\sum_{a,i} (\theta_{t,a,i} * K_{a,i})}{\sum_{a,i} (K_{a,i})}$ (7)	*Average site and portfolio capacity factor* (per time step, averaged over all sites and technologies)

[2] The capacity factors presented in Table 1 are calculated as weighted arithmetic mean to consider the different relevance of the infeed profiles. The capacity weighting yet necessitates a backup rule to cope with sites where all capacities are zero. We hereby use the average capacity shares $\bar{k}_{\cdot,i} = \frac{\sum_a K_{a,i}}{\sum_{a,i'} K_{a,i'}}$ in the overall capacity mix. These capacity shares $\bar{k}_{\cdot,i}$ are then used in Eqs. (4) and (5) whenever the sum of capacities $\sum_i (K_{a,i})$ at one site a is equal to zero.

$\bar{\theta}_{\cdot,\cdot,\cdot} = \frac{\sum_t \bar{\theta}_{t,\cdot,\cdot}}{N_T} = \frac{\sum_{t,a,i}(\theta_{t,a,i} * K_{a,i})}{N_T \sum_{a,i}(K_{a,i})}$ (8)	*Overall capacity factor* (averaged over time as well as all sites and technologies)
$c_{\cdot}^{ann} = \frac{\sum_{a,i}(c_i^{ann} K_{a,i})}{\sum_{a,i}(K_{a,i})}$ (9)	Average cost of the portfolio[3] cost component averaged over the whole portfolio.

In the first step of a number of mathematical reformulations, the right-hand side expression is extended by the *average time and portfolio capacity factor* $\bar{\theta}_{\cdot,a,\cdot}$ averaged over time and technologies (cf. Eq. (10)) to allow the analysis of deviations between the individual capacity factor and its mean. This step corresponds to adding and subtracting simultaneously the term $\bar{\theta}_{\cdot,a,\cdot} K_{a,i}$ which does not change the result. A similar approach is used several times in the following.

$$Y_{t,a,i} = \bar{\theta}_{\cdot,a,\cdot} K_{a,i} + (\theta_{t,a,i} - \bar{\theta}_{\cdot,a,\cdot}) K_{a,i} \tag{10}$$

After expanding Eq. (10) further by inserting the two mathematically neutral terms $\frac{\bar{\theta}_{t,\cdot,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \bar{\theta}_{\cdot,a,\cdot} - \frac{\bar{\theta}_{t,\cdot,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \bar{\theta}_{\cdot,a,\cdot}$ and $\frac{\bar{\theta}_{t,a,\cdot}}{\bar{\theta}_{\cdot,a,\cdot}} \bar{\theta}_{\cdot,a,\cdot} - \frac{\bar{\theta}_{t,a,\cdot}}{\bar{\theta}_{\cdot,a,\cdot}} \bar{\theta}_{\cdot,a,\cdot}$, some further reformulations lead to the expression given in Eq. (11).

$$Y_{t,a,i} = \left(\bar{\theta}_{\cdot,a,\cdot} + \frac{\bar{\theta}_{\cdot,a,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} (\bar{\theta}_{t,\cdot,\cdot} - \bar{\theta}_{\cdot,\cdot,\cdot}) + \bar{\theta}_{\cdot,a,\cdot} \left(\frac{\bar{\theta}_{t,a,\cdot}}{\bar{\theta}_{\cdot,a,\cdot}} - \frac{\bar{\theta}_{t,\cdot,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \right) + (\theta_{t,a,i} - \bar{\theta}_{t,a,\cdot}) \right) K_{a,i} \tag{11}$$

Insertion of Eq. (11) in Eq. (2) and decomposition of the cost term similarly to Eq. (10) along with rearranging leads to the following expression for the annualized value (subsequently also denoted as profit):

$$\begin{aligned} AV_{a,i} = \Bigg(& \sum_t \bar{\theta}_{\cdot,a,\cdot} p_t - c_{\cdot} + \sum_t \frac{\bar{\theta}_{\cdot,a,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} (\bar{\theta}_{t,\cdot,\cdot} - \bar{\theta}_{\cdot,\cdot,\cdot}) p_t \\ & + \sum_t \bar{\theta}_{\cdot,a,\cdot} \left(\frac{\theta_{t,a,\cdot}}{\bar{\theta}_{\cdot,a,\cdot}} - \frac{\bar{\theta}_{t,\cdot,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \right) p_t + \sum_t (\theta_{t,a,i} - \bar{\theta}_{t,a,\cdot}) p_t \\ & - (c_i^{ann} - c_{\cdot}^{ann}) \Bigg) K_{a,i} \end{aligned} \tag{12}$$

Based on Eq. (12), four different value components may be identified. These are defined as capacity specific values and labeled *yield, resource, site and technology component*. In the following, the different value components are characterized for the case of zonal price time series

[3] Note that we assume the same technology specific cost for all sites. Notably if offshore wind sites would be included in the analysis, a differentiation between sites would be necessary.

(cf. Eq. (13) to (17)). To consider scenarios with nodal prices and accordingly the influence of e.g. grid congestion on aggregation, an extended version of Eq. (12) is presented in Appendix A. Related adjustments and extensions of the value components are discussed accordingly.

The *yield component* given in Eq. (13) corresponds to the average value of all technologies installed at one site a and its interpretation is rather straightforward. It is driven by the time and the a*verage time and portfolio capacity factor* $\bar{\theta}_{\cdot,a,\cdot}$ shown in Table 1 and is thus directly related to the renewable infeed i.e. the full load hours $FLH = \bar{\theta}_{\cdot,a,\cdot} N_T \Delta t$ at each site (with $N_T \Delta t$ the length of the representative period).

$$VC_{a,i}^{yield} = \sum_t \bar{\theta}_{\cdot,a,\cdot} p_t - c_{\cdot}^{ann} = \bar{\theta}_{\cdot,a,\cdot} \cdot N_T \Delta t \cdot p_{\cdot} - c_{\cdot}^{ann} \tag{13}$$

The yield component is calculated as difference between the revenue at site a at average prices $p_{\cdot}$ and the average (site-independent) cost $c_{\cdot}$ of the portfolio. Hence, there is no differentiation in the yield component between technologies at one site. The yield component only differs for investment choices at different locations, i.e. with different infeed profiles. It contributes positively to the value of an investment choice as long as the revenue within the representative period of the average portfolio at one site is greater than the average cost of the portfolio.

The *resource component* as given in Eq. (14) is driven by the market value factor m of the renewable technology (cf. Hirth (2013)) and thus indicates the self-destructive effect of rising renewable shares in electricity markets. This effect results from negative correlations between electricity prices p_t and renewable infeed (represented by the a*verage site and portfolio capacity factor* $\bar{\theta}_{t,\cdot,\cdot}$), as indicated by the second equality in Eq. (14).

$$\begin{aligned} VC_{a,i}^{resource} &= \sum_t \frac{\bar{\theta}_{\cdot,a,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \left(\bar{\theta}_{t,\cdot,\cdot} - \bar{\theta}_{\cdot,\cdot,\cdot}\right) p_t = \frac{\bar{\theta}_{\cdot,a,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \sum_t \left(\bar{\theta}_{t,\cdot,\cdot} - \bar{\theta}_{\cdot,\cdot,\cdot}\right)(p_t - p_{\cdot}) \\ &= \bar{\theta}_{\cdot,a,\cdot} \sum_t \left(\frac{\bar{\theta}_{t,\cdot,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} - 1\right) p_t = \bar{\theta}_{\cdot,a,\cdot} (m-1) \cdot N_T \Delta t \cdot p_{\cdot} \end{aligned} \tag{14}$$

The *site component* describes the influence of local time-specific infeed profiles on the value of an investment choice. As shown in Eq. (15) it is driven by the difference of the capacity factor weighted over technologies and the capacity factor weighted over sites and technologies.

$$VC_{a,i}^{site} = \sum_t \bar{\theta}_{\cdot,a,\cdot} \left(\frac{\theta_{t,a,\cdot}}{\bar{\theta}_{\cdot,a,\cdot}} - \frac{\bar{\theta}_{t,\cdot,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}}\right) p_t = \bar{\theta}_{\cdot,a,\cdot} \sum_t \left(\frac{\theta_{t,a,\cdot} - \bar{\theta}_{\cdot,a,\cdot}}{\bar{\theta}_{\cdot,a,\cdot}} - \frac{\bar{\theta}_{t,\cdot,\cdot} - \bar{\theta}_{\cdot,\cdot,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}}\right)(p_t - p_{\cdot}) \tag{15}$$

As indicated by the second equality in Eq. (15), the site-specific component is positive, if the (relative) site-specific capacity factor is more positively (or rather: less negatively) correlated with the price variations than the average capacity factor.

The *technology component* as shown in Eq. (16) finally indicates the impact of the technology type i on the value of an investment choice. It is driven by the difference between the technology-specific and the average portfolio capacity factor per site and the cost deviation from the average portfolio cost.

$$VC_{a,i}^{technology} = \sum_{t} (\theta_{t,a,i} - \bar{\theta}_{t,a,\cdot}) p_t - (c_i^{ann} - c_{\cdot}^{ann}) \tag{16}$$

A technology with relatively high capacity factors (compared to the average) is able to compensate relatively high costs and hence the corresponding technology component is positive. Conversely a technology with relatively low capacity factors will show a positive technology component if low capital expenditures overcompensate for the losses in revenues. The first case is yet limited in case there are many hours with low or negative prices caused by high renewable infeed – then these hours cannot yield a positive contribution to the technology component.

Subsequently, we merge the yield and the resource component in view of a more compact presentation. This leads to the net-yield component corresponding to the sum of yield and resource component as shown in Eq. (17). Note that each component is considered independently for the clustering even though they are not presented individually in the results.

$$VC_{a,i}^{net\ yield} = VC_{a,i}^{yield} + VC_{a,i}^{resource} = \sum_{t} \frac{\bar{\theta}_{t,\cdot,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \bar{\theta}_{\cdot,a,\cdot} p_t - c_{\cdot}^{ann} \tag{17}$$

3.2.2 Illustrative example

Figure 2 shows the value of four exemplary investment choices based on the *Scenario Future* discussed in detail in chapter 4. These investment choices correspond to two different wind turbine types placed in one northern site (Nuts3 region DEF02) and one southern site (Nuts3 region DE132) in Germany. One turbine type for low wind speed (type 2) and one for high wind speed (type 8) are chosen to illustrate the varying decomposition of the profit.

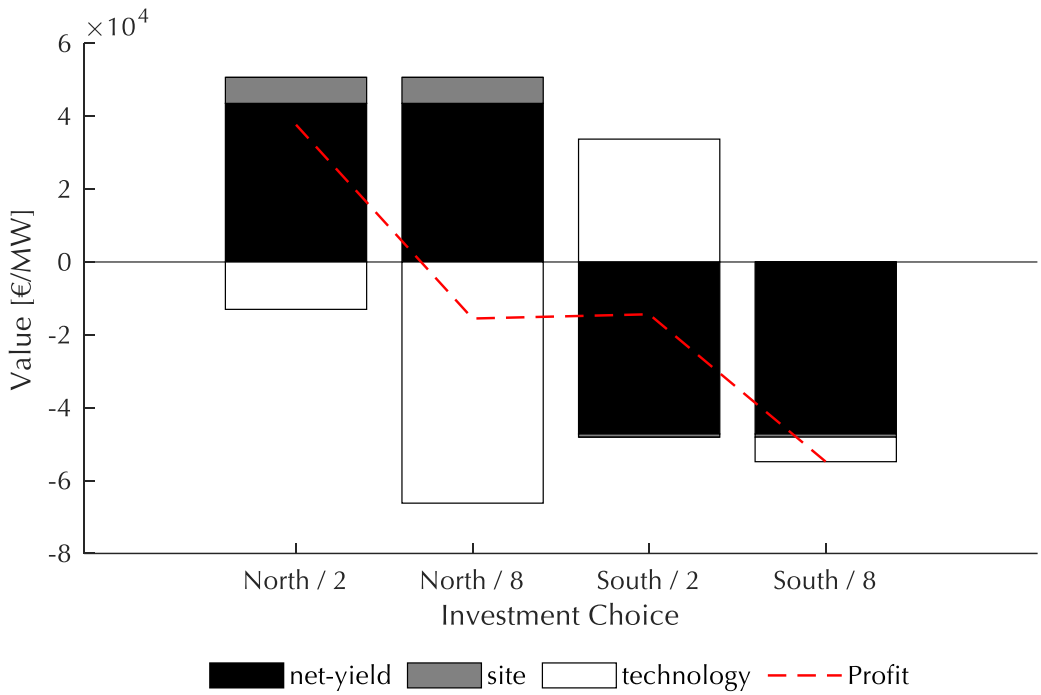

Figure 2: Decomposed value of four investment choices of the *Scenario Future* (cf. section 4)

By definition, the net-yield component and the site component of different technologies at the same site are equal (cf. Figure 2 as well as Eqs. (15) and (17)). The values yet significantly differ between locations. The higher value of the net-yield component in the north reflects the major difference in full load hours between north and south Germany. The site component contributes positively in the north and only slightly negatively at the southern site in this example. The technology component varies between all investment choices. As stated in Eq. (16), there is one main aspect to keep in mind for this component. The technology component can only be positive if the summed difference between the technology-specific and the portfolio-average capacity factor weighted by the electricity price is larger than the corresponding difference in investment cost.

3.3 Scenario definition

The selection of representative scenarios is crucial for the robustness of the clustering results and hence the selected scenarios should be sufficiently contrasted. Put differently: the more fundamental the differences between the scenarios, the less sensitive are the resulting conclusions to changes in input parameters and consequently the identified aggregated investment choices may be applied to quite different energy systems and optimization models.

As presented in section 3.2.1, capacity factors, installed capacities, prices and investment costs are the key parameters necessary to calculate the value components. The capacity factors are based on technology data like e.g. hub height, rated power and rotor diameter of a wind turbine

and weather data like wind velocities. The latter are dependent on the selected representative period, whereas installed capacities and prices are scenario-dependent[4].

3.4 Clustering algorithm and parameter settings

A large number of cluster algorithms has been developed over the last decades and may be selected to perform the intended aggregation of investment choices. A detailed classification of clustering algorithms and a summary of the underlying theory is provided e.g. by Gan et al. (2007) whereas Aghabozorgi et al. (2015) present a review of clustering approaches for time-series. In the following a short overview and justification of the applied algorithms is given.

The clustering of investment choices can be carried out by a relatively simple partitioning algorithm and does not require any special enhancement. The approach at hand utilizes the k-means algorithm which currently is one of the mostly used and best bench-marked clustering algorithms (cf. Saxena et al. (2017)). Figure 3 (right) shows the flow chart of the utilized k-means algorithm. The investment choices (IC) represent the objects and the value components represent the attributes for the clustering. The centroids are calculated as average of the data points within each cluster. Furthermore, the data points are assigned to clusters by minimizing the squared-Euclidean distance between the objects (all IC) and the centroids (representative IC) of each cluster. Despite the advantages of k-means there are also several drawbacks, e.g. the sensitivity to outliers and noise as well as the importance of the initialization i.e. selecting good initial clusters (cf. Gan et al., 2007). The number of clusters into which the investment choices will be assigned can be set randomly by the user or by applying a heuristic and the initial clusters (seeds) can be set randomly or systematically as well.

[4] Note that the technology data and the geographical coverage and resolution must not vary between the scenarios to avoid diverging investment choices between the scenarios. In that case the clustering will not provide consistent results. Also the meteorological year(s) used as representative period should be the same for all scenarios to avoid bias by random volatilities in weather data.

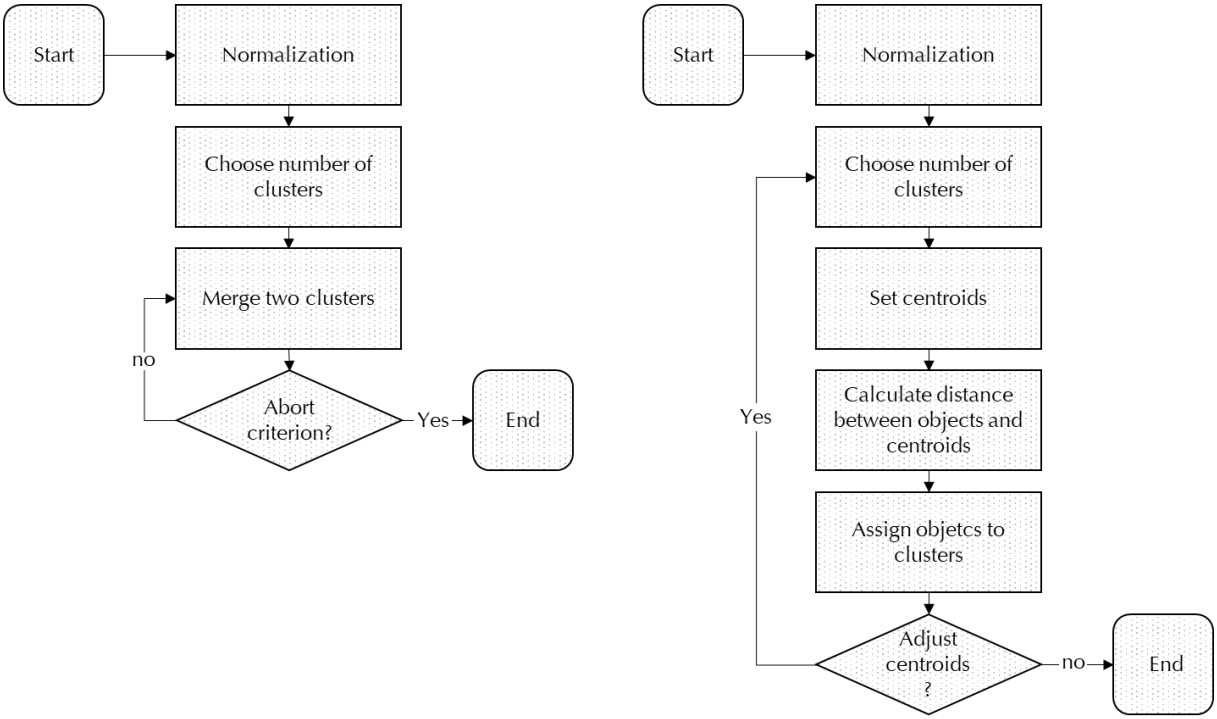

Figure 3: Flow diagram of hierarchical (left) and k-means (right) cluster algorithms

To counter the drawbacks of k-means, a suitable cluster number and seeds are obtained by applying an agglomerative hierarchical cluster algorithm (cf. flow chart in Figure 3 left) as preprocessing step (cf. Bacher et al., 2010; Fisher, 1996; Milligan, 1980). Agglomerative hierarchical clustering means that starting with every single object (i.e. IC) representing one separate cluster, the number of clusters is reduced iteratively by merging two clusters into a new one. The clustering ends when all objects are merged within one cluster or when some other abortion criterion is satisfied. The algorithm applied utilizes the change in within-cluster sum of squared Euclidean distances for each merging step as this measure assesses the quality of the newly build cluster formation pretty well.

4 Application – The case of wind energy in Germany

In this chapter, the proposed approach is applied to a German case study. Germany encompasses quite diverse geographical conditions including coastal regions as well as inland regions with mountain sites and forest areas. The application is focused on wind power – as the analysis of wind energy expansion is expected to be the most complex case of renewable energy. Moreover, less portfolio effects are expected for e.g. photovoltaic since its technological diversity is not as high as for wind – at least for any given site (i.e. large-scale ground-mounted and roof-top modules are placed in different sites). Anyhow, the algorithm can be applied to other spatial resolutions and technologies. The investment choices based on German data are defined in section 4.1 and three distinct scenarios are developed in section 4.2. In section 4.3 respectively 4.4, the value components are calculated and clustered.

4.1 Step 1: Define investment choices as objects for the clustering algorithm

Per definition the number of investment choices (combination of site and technology type) is directly related to the spatial resolution and level of technological detail considered. Furthermore, the set of sites and technology types defines the size of the clustering problem. In the following, the selection of sites and wind turbine types is set out.

A rather high spatial resolution is selected by distinguishing the 402 German NUTS3 regions (cf. EU (2013)[5]). These correspond to counties ("Landkreise") and non-county cities ("kreisfreie Städte") and for these mostly individual wind velocities[6] are available. The sites are adequately defined by their coordinates and related wind velocities.

Eight different types of wind turbines are selected to represent technology choices available in Germany. Table 2 presents the utilized wind turbines and lists the parameters required for further calculations. The selection of wind turbines is thereby two-fold: On the one hand, a k-medoids cluster analysis of the existing German wind power plant fleet in 2017 is conducted. The wind turbines 1 to 5 and 8 are representatives of the latter grouping. On the other hand, technological developments are considered by the selection of two additional turbines types, named turbine 6 and 7. Both turbines are inspired by combined technology projections of Deutsche Windguard (2015), Bundesministerium für Wirtschaft und Energie (2014) and own assumptions. The

[5] For reasons of internal model compatibility, the NUTS3 edition 2013 instead of the most recent edition was utilized in our analysis. Since the differences only involve four districts and minor boundary shifts, they do not affect our analysis.

[6] The wind velocities are publicly available and downloaded from the ftp server of the German Meteorological Service DWD (ftp://opendata.dwd.de/climate_environment/CDC/).

consideration of technological development and its representation within the investment choices is crucial when optimizing the capacity expansion of wind power plants in any future scenario.

Table 2: Technology data of the eight representative onshore wind turbines

Turbine Type	Hub height [m]	Rotor Diameter [m]	Power [kW]	Type[7]	Capex [€/kW]
1	72	53	800	High speed	1.047
2	139	121	2.530	Low speed	1.571
3	109	92	2.350	High speed	1.155
4	142	114	3.170	Low speed	1.290
5	110	109	3.000	Low speed	1.169
6	150	140	4.000	Low speed	1.573
7	120	124	4.500	High speed	1.363
8	120	140	6.000	High speed	1.483

Hub height, rotor diameter, power and wind velocities are used to calculate the infeed characteristics of each turbine at every site. The investment costs enter the yield and the technology component as defined in section 3.2.1

Since the literature does not provide cost data for turbine 1 to 8, the costs presented in Table 2 have been derived based on Fingersh et al. (2006). The average currency exchange rate of 2006 (1.26$/€ cf. Bundesbank, 2020) enters the currency conversion and an adjustment factor of 1.2 is applied to fit the results to comparable turbine costs in the literature (cf. Bundesministerium für Wirtschaft und Energie, 2014; Deutsche Windguard, 2015).

Since the geographical setting is similar for all scenarios, only the capacities need to be adapted. A simple clustering algorithm with predefined centroids is utilized to allocate the wind capacities of each scenario to the eight representative technology types per site. In total 402 sites and 8 technology types lead to a total of 3216 investment choices i.e. cluster objects.

[7] Low- and high-speed wind turbines are differentiated by their specific power (power per swept area), whereas turbines with a specific power of more than 0.35 kW/m² are defined as high speed turbines.

4.2 Step 2: Choice of adequate scenarios

Three distinct scenarios are chosen to calculate the value components and as discussed in section 3.3, each scenario is primarily defined by electricity prices and installed capacities.

Scenario 2017	→ Actual capacities and prices of 2017
Scenario Future	→ Optimized capacities and prices for a future year
Scenario Nodal	→ Capacities (2017) scaled and approximated prices for 2020

Scenario 2017 represents a real data scenario and can be interpreted as "status quo scenario". The capacities match the size of the real wind power installations in Germany 2017 and the prices are set to the historical day-ahead spot prices.

The prices and capacities for the *Scenario Future* are the result of a simplified optimizing electricity market model. The problem is designed as greenfield approach with one conventional backup technology and three renewable sources, namely PV, wind offshore and wind onshore. It is constructed as a low-emission scenario with a CO2 budget of 65 Mio. tons per year for Germany which leads to a renewable share of about 65 % of the overall production. The demand[8] is held constant from 2017 in expectation of balancing effects between increasing energy efficiency and growth in population and technology usage. The mathematical description of the underlying optimization problem is provided in Appendix B.

The third scenario, *Scenario Nodal*, is designed to ensure the consideration of congestion effects in the transmission grid (cf. Appendix A). By utilizing nodal prices, the influence of over-supply and supply shortages caused e.g. by concentration of renewables in the north of Germany is considered. Thereby, this scenario introduces further systemic aspects in the developed algorithm, namely a grid value component (cf. Appendix A). The scenario is adopted from the work of Felling and Weber (2018). The prices are calculated as average timeseries of the six presented scenarios in the latter work, whereas the distribution of capacities is taken from the actual capacities in 2017 and scaled to the overall capacity in 2020 assumed by Felling and Weber (2018).

4.3 Step 3: Calculate value components for various scenarios

For the investment choices defined in step 2, the value components are computed according to the methodology described in section 3.2. Based on these calculations, Figure 4 presents the 50 most profitable investment choices of each scenario in descending order. Various structural

[8] The demand is determined by utilization of the demand profile for Germany available at the ENTSOE Transparency Platform (2019) scaled by the final consumption of Germany in 2017 (521 TWh cf. International Energy Agency (2018)) and with missing data filled through interpolation.

differences between the scenarios as well as between the investment choices within each scenario are observable and will be discussed in the following.

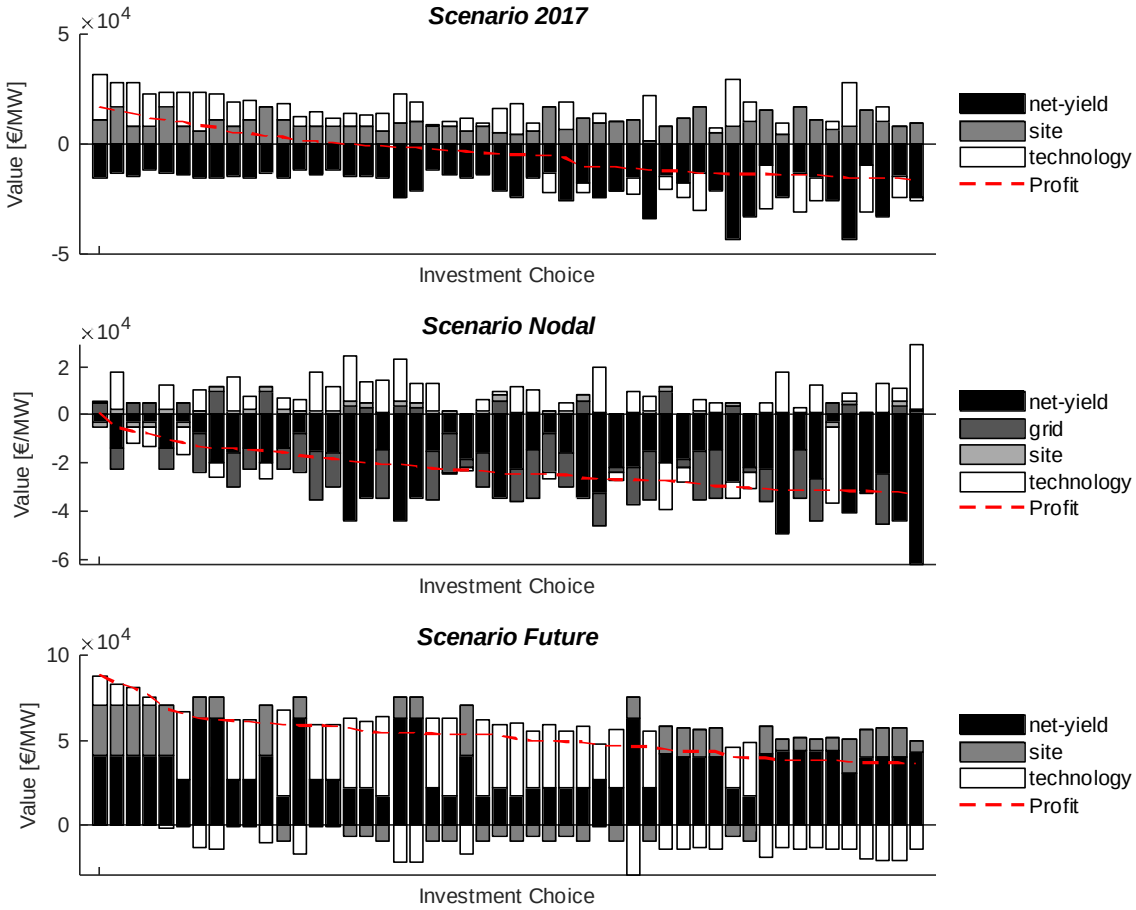

Figure 4: Value components of the 50 most profitable investment choices per scenario

The number of investment choices with positive profits ranges from 15 in the *Scenario 2017* over 1 in the *Scenario Nodal* to 191 in the *Scenario Future*. Apparently, these numbers run against the number of realized investment choices i.e. with a capacity greater zero (991, 927 and 54). This major shift between profitable investment choices and those with capacities greater zero between the two first scenarios and the last scenario demonstrate fundamental difference in the underlying data: The great number of realized wind sites in 2017 and in a nodal case (based on 2017) are influenced by the fixed feed in tariffs paid in Germany over the last decades. Furthermore, the price increase in the *Scenario Future* is based on the restriction of CO2 emissions and the absence of subsidies within the optimization framework. Consequently, the number of profitable investment choices at wholesale market prices increases. The fact that only a fraction of the profitable investment choices in the *Scenario Future* are realized is explainable by the maximum capacity restriction on each site. It is defined on a per site basis and not individually for each investment choice. Therefore, the technology with the highest profit is built up to the site capacity

limit and other profitable turbine types are disregarded. Moreover, a site where no technology has positive profit remains unused. This implies that the number of realized investment choices depends on the maximum capacity installable at the most profitable sites.

In the *Scenario Nodal*, a reduction in overall profitability, a decrease of the site component and in many cases negative contributions of the grid component are furthermore observed. Main driver for these observations is the change in prices due to the consideration of grid restrictions.

4.4 Step 4: Clustering of value components

In the following two sub-sections 4.4.1 and 4.4.2 the two clustering algorithms are applied. As described in section 4.2, the cluster analysis is based on 3216 objects representing 3216 investment choices and 13 value components as attributes (three scenarios with 4 respectively 5 value components).

4.4.1 Predefine cluster numbers using hierarchical clustering

To identify an adequate number of clusters as starting point for the k-means clustering algorithm, an agglomerative hierarchical clustering approach is applied. The resulting hierarchy is visualized as dendrogram in Figure 5 (right) with the last 11 clusters highlighted in colors. Furthermore, the Within- and Between-Cluster sum of squared distances of the hierarchical clustering in comparison to the objective function value of the k-means clustering are shown in Figure 5 (left). It can be observed that the sum of distances within the clusters rises significantly when the number of clusters decreases below (approximately) 11. Moreover, the advantage of k-means clustering is recognizable by the lower sum of squared distances compared to the hierarchical pre-clustering.

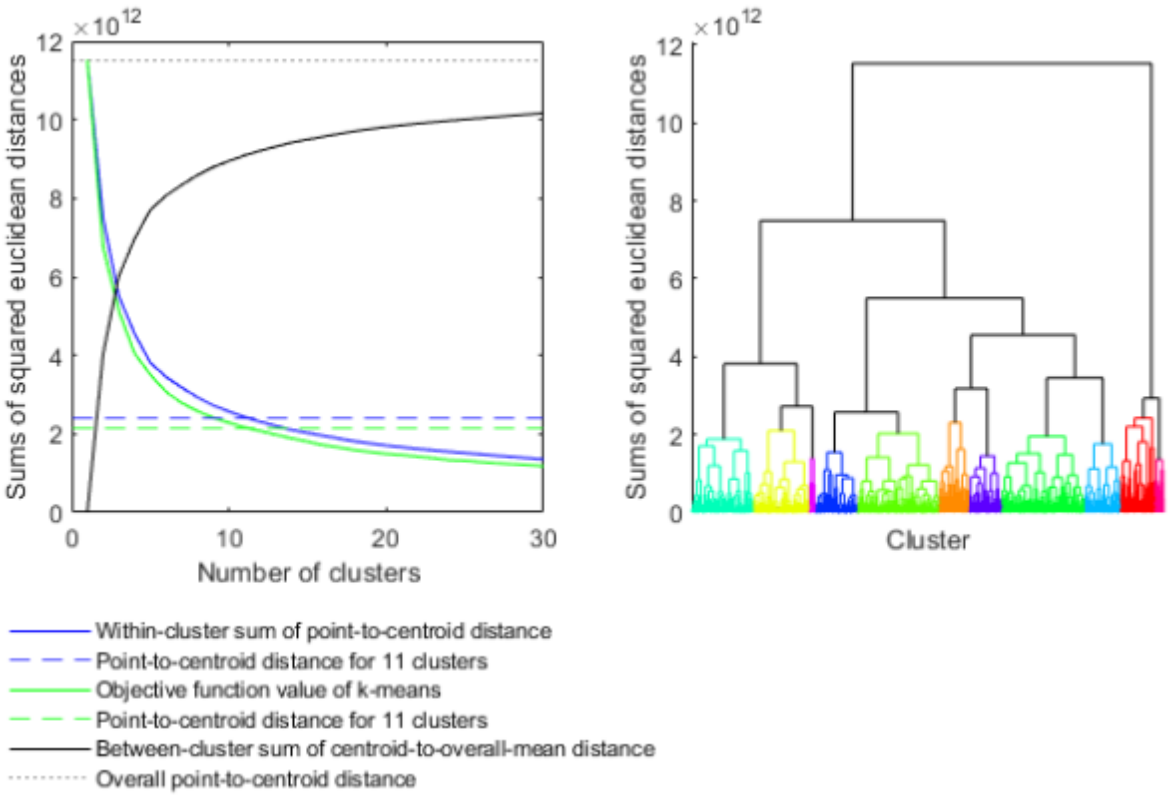

Figure 5: Sums of squared Euclidean distances of the hierarchical cluster algorithm and k-means

4.4.2 Aggregation of investment choices

The following analysis focusses on two key aspects in the assignment of investment choices to clusters. The first more intuitive one is the geographical and technological diversification of clusters and investment choices. The second aspect are the characteristics of cluster specific value components.

Figure 6 presents the geographic position of three exemplary cluster centroids (averaged site coordinates[9] of all objects within the cluster) and the corresponding investment choices.

[9] Note that averaging the objects within each cluster might be appropriate for geographical aspects but is not reasonable for technological characteristics.

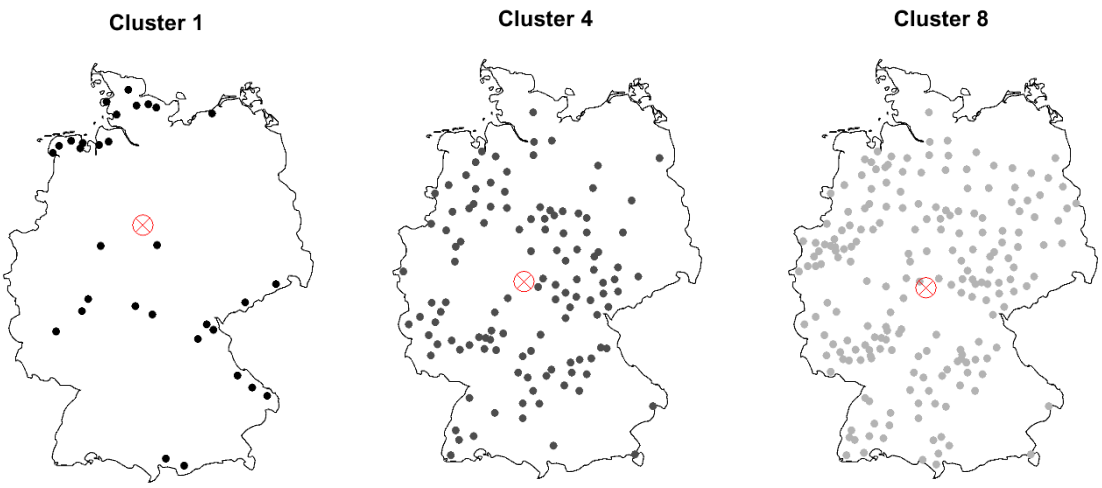

Figure 6: Geographical representation of three different clusters (centroid in red)

The representation of single clusters in Figure 6 as well as the position of all 11 centroids shown in Figure 7 indicate two distinct types of clusters. On the one hand, there are some clusters with only few profitable investment choices, including many sites in northern Germany (cf. cluster 1 in Figure 6). On the other hand, there are various clusters gathering multiple less profitable investment choices, whose centroids are concentrated in central Germany. The size of the bubbles in Figure 7 is related to the cumulated wind capacity potential of the investment choices in each cluster. The potential of each site is only considered once by assigning it to the cluster with the highest profitability that includes this site.

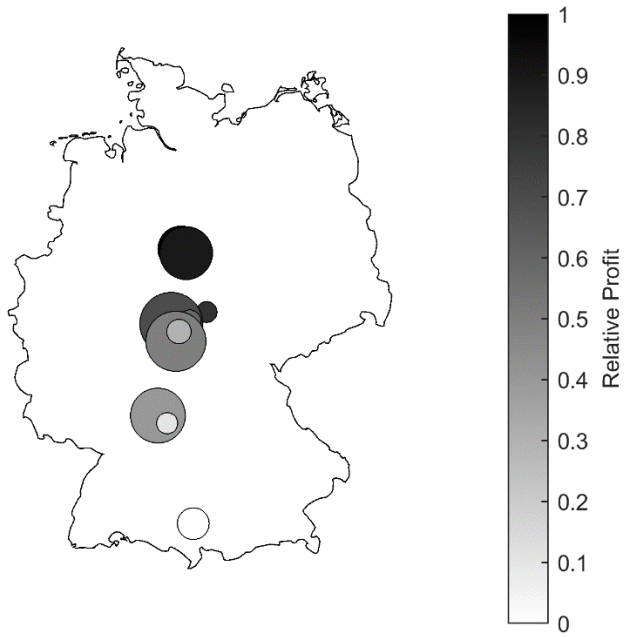

Figure 7: Geographical representation of all 11 cluster centroids

The distribution of cluster centroids in Figure 7 reveals an accumulation along the north-south axis, which implies that there is no strict distinction of east and west in the clustering. This finding

seems plausible for Germany as the suitability of sites for wind generation and the actual distribution of high and low wind turbines is changing rather from north to south than from west to east.

The technology distribution within the clusters is presented in Figure 8, with clusters ordered by decreasing overall profitability (cf. Figure 9). It supports the conclusion that only few investment choices are present in the most profitable clusters. Interesting is the fact, that the distribution of turbine types is rather similar for three of the four most profitable clusters, with types 7 and 8 being very rarely used. Clusters 3, 5, 9 and 10 are quite specialized on high speed wind turbines, but the low profitability of clusters 9 and 10 and the small number of units in clusters 3 and 5 support the idea that technological diversification does not contribute much to the overall profitability.

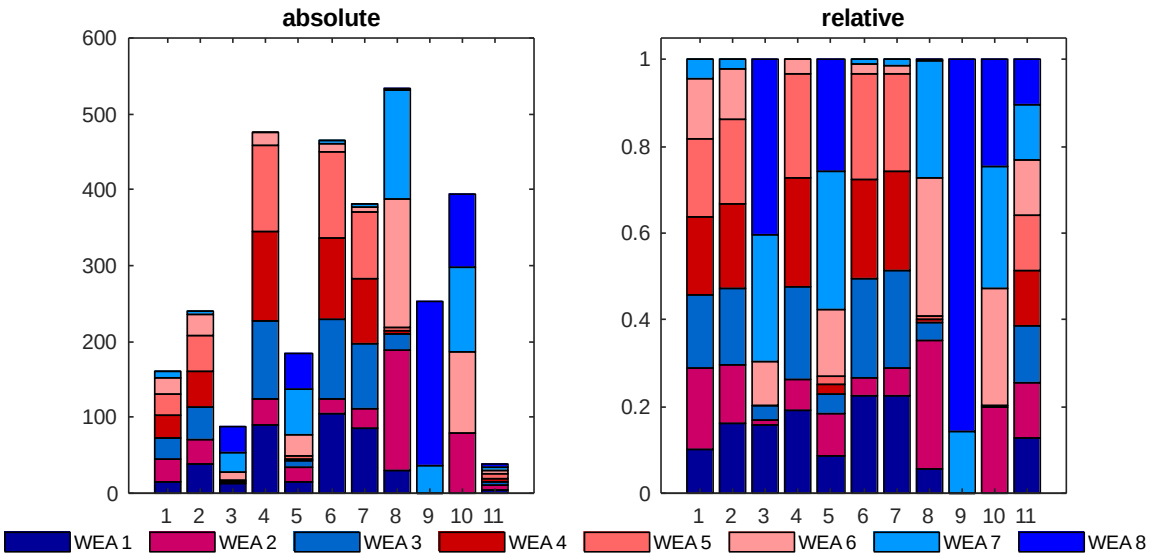

Figure 8: Distribution of technology types within clusters (high speed turbines in blue and low speed turbines in red)

In the following, the clusters are analyzed based on the cluster specific value components. Figure 9 presents the profit per cluster in descending order.

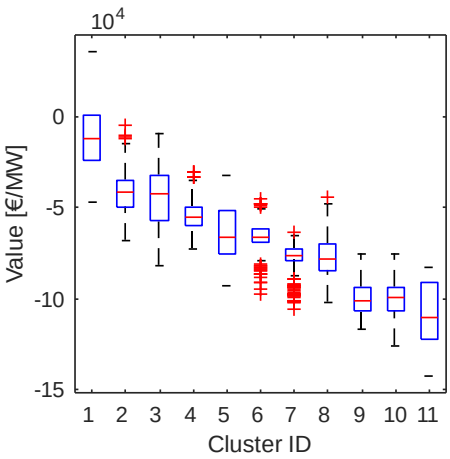

Figure 9: Average profit of all scenarios in descending order

The profit under the different scenarios is displayed in Figure 10 side by side. A similar ranking of clusters is found in all scenarios, this indicates that the ranking of sites remains rather constant under different market regimes.

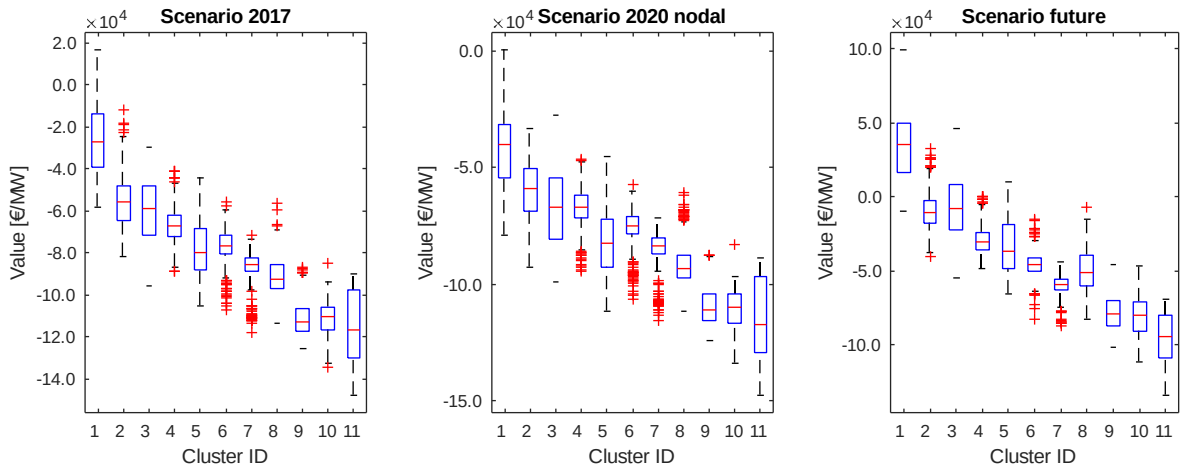

Figure 10: Profit per scenario in descending order

Figure 11 presents the main value components averaged over the three scenarios. The variation between clusters is highest in absolute terms for the technology and the net-yield component. The components are strongly varying even between neighboring clusters, yet some balancing effect may be observed when it comes to overall profitability. E.g. clusters 2 and 3 are rather similar in overall profitability (cf. Figure 10) yet cluster 2 has a much more positive technology component whereas cluster 3 has a higher net yield. Given the high influence of the net-yield component it is not that surprising that this cluster value component is often aligned with the overall profit. In absolute value, the site component has the smallest impact. Apart from that, the variability is not that high, only the values for cluster 1 and 3 are a little higher. Regarding the grid component, three clusters single out. Cluster one is the only one with a strongly negative grid component This is related to the negative influence of the grid component in times of high

renewable infeed (cf. Appendix A). Because of scarce transfer capacities in times of high renewable infeed, the grid component becomes negative. This especially influences cluster 1 – accordingly this cluster includes mostly investment choices in the north, where transmission capacities get scarce first. The median grid component of the other investment choices is approximately zero with some substantial negative outliers. On the other hand, clusters 1 to 3 show the by far largest interquartile range without significant outliers.

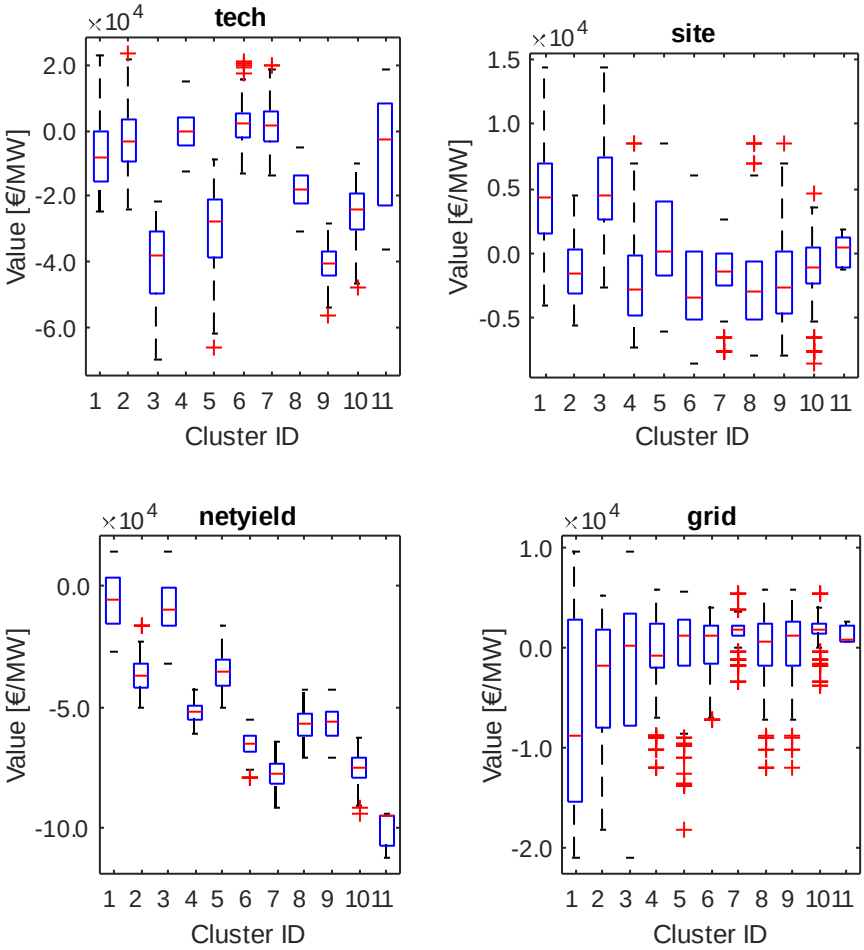

Figure 11: Characteristics of the four value components

5 Conclusions and outlook

The paper at hand proposes a new aggregation approach to be used in view of large-scale electricity markets. It is based on the clustering of various value components, related e.g. to the total yield or the site-specific infeed profile of investment choices. A case study for the German electricity system is applied and the influence of various value components on the combined technological and spatial aggregation is analyzed.

Four main insights for the modeling of electricity markets and the related political framework may be retained:

1. Clustering the investments choices as described allows to reduce the quantity of investment choices that have to be included in long-term electricity market models without compromising the quality of results. Clear differences in profitability can be identified for different investment choices which justify their grouping in a limited number of clusters.
2. Furthermore, we show that aggregation based on economic parameters complements the standard methods of aggregation based on time-series clustering. Our approach helps closing the gap in spatial and technological aggregation, yet a comparison of the results of our approach to other approaches is left for further work.
3. The results suggest that the spatial heterogeneity, i.e. the diversity of wind profiles at different sites, does not contribute much to the investment value. Meaning that complementarity in sites is not that important for overall optimal system designs. This result supports the recent work of inter alia Hirth and Müller (2016) and Klie and Madlener (2020).
4. The results based on three distinct scenarios emphasize the influence of political framework conditions and market prices on the valuation of investment choices. On the one hand, the relatively high diversification of realized investment choices in the *Scenario 2017* compared to the *Scenario Future* shows the influence of support systems like the EEG. On the other hand, capacity expansion without further restrictions, i.e. under a "copperplate" assumption, can obviously lead to strong effects of capacity concentration at highly profitable sites.

In a next step, the developed methodology should be applied to input data for a large electricity market model and the impacts regarding computational acceleration and the aggregation-related error should be quantified.

Acknowledgement

We gratefully acknowledge the financial support of the German Ministry of Economics and Technology (BMWi) and the project supervision of Projektträger Jülich (PtJ) within the project WeatherAggReOpt-Time (grant number 03ET4042A) for enabling the underlying research described in this paper.

References

Adhau, S.P., Moharil, R.M., Adhau, P.G., 2014. K-Means clustering technique applied to availability of micro hydro power. Sustainable Energy Technologies and Assessments 8, 191–201. https://doi.org/10.1016/j.seta.2014.09.001.

Aghabozorgi, S., Seyed Shirkhorshidi, A., Ying Wah, T., 2015. Time-series clustering – A decade review. Information Systems 53, 16–38. https://doi.org/10.1016/j.is.2015.04.007.

Ali, M., Ilie, I.-S., Milanovic, J.V., Chicco, G., 2013. Wind Farm Model Aggregation Using Probabilistic Clustering. IEEE Trans. Power Syst. 28 (1), 309–316. https://doi.org/10.1109/TPWRS.2012.2204282.

Bacher, J., Pöge, A., Wenzig, K., 2010. Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren, 3rd ed. Oldenbourg, München, 538 pp.

Bahl, B., Kümpel, A., Seele, H., Lampe, M., Bardow, A., 2017. Time-series aggregation for synthesis problems by bounding error in the objective function. Energy 135, 900–912. https://doi.org/10.1016/j.energy.2017.06.082.

Bloom, J.A., 1983. Solving an Electricity Generating Capacity Expansion Planning Problem by Generalized Benders' Decomposition. Operations Research 31 (1), 84–100. https://doi.org/10.1287/opre.31.1.84.

Breuer, C., Moser, A., 2014. Optimized bidding area delimitations and their impact on electricity markets and congestion management, in: 11th International Conference 2014, pp. 1–5.

Bucksteeg, M., 2019. Modelling the impact of geographical diversification of wind turbines on the required firm capacity in Germany. Applied Energy 235, 1476–1491. https://doi.org/10.1016/j.apenergy.2018.11.031.

Bundesbank, D., 2020. Wechselkursstatistik, 13.

Burstedde, B., 2012. From nodal to zonal pricing: A bottom-up approach to the second-best, in: 9th International Conference 2012, pp. 1–8.

Deutsche Windguard, 2015. Kostensituation der Windenergie an Land in Deutschland - Update SP15027A1, Varel, 65 pp. https://www.windguard.de/veroeffentlichungen.html?file=files/cto_layout/img/unternehmen/veroeffentlichungen/2015/Kostensituation%20der%20Windenergie%20an%20Land%20in%20Deutschland%20-%20Update.pdf (accessed 14 August 2019).

Domínguez-Muñoz, F., Cejudo-López, J.M., Carrillo-Andrés, A., Gallardo-Salazar, M., 2011. Selection of typical demand days for CHP optimization. Energy and Buildings 43 (11), 3036–3043. https://doi.org/10.1016/j.enbuild.2011.07.024.

ENTSOE Transparency Platform, 2019. https://transparency.entsoe.eu/ (accessed 16 August 2019).

EU, 2013. Verordnung (EU) Nr. 1319/2013 der Kommission vom 9. Dezember 2013 zur Änderung der Anhänge der Verordnung (EG) Nr. 1059/2003 des Europäischen Parlaments und des Rates über die Schaffung einer gemeinsamen Klassifikation der Gebietseinheiten für die Statistik (NUTS), 57 pp.

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

Felling, T., Weber, C., 2018. Consistent and robust delimitation of price zones under uncertainty with an application to Central Western Europe. Energy Economics 75, 583–601. https://doi.org/10.1016/j.eneco.2018.09.012.

Fingersh, L., Hand, M., Laxson, A., 2006. Wind Turbine Design Cost and Scaling Model. Technical Report NREL/TP-500-40566. https://www.nrel.gov/docs/fy07osti/40566.pdf (accessed 14 August 2019).

Fisher, D., 1996. Iterative Optimization and Simplification of Hierarchical Clusterings. jair 4, 147–178. https://doi.org/10.1613/jair.276.

Frew, B.A., Becker, S., Dvorak, M.J., Andresen, G.B., Jacobson, M.Z., 2016. Flexibility mechanisms and pathways to a highly renewable US electricity future. Energy 101, 65–78. https://doi.org/10.1016/j.energy.2016.01.079.

Gan, G., Ma, C., Wu, J., 2007. Data Clustering: Theory, Algorithms, and Applications. Society for Industrial and Applied Mathematics.

Gómez-Muñoz, V.M., Porta-Gándara, M.A., 2002. Local wind patterns for modeling renewable energy systems by means of cluster analysis techniques. Renewable Energy 25 (2), 171–182. https://doi.org/10.1016/S0960-1481(01)00013-1.

Green, R., Staffell, I., Vasilakos, N., 2014. Divide and Conquer k-Means Clustering of Demand Data Allows Rapid and Accurate Simulations of the British Electricity System. IEEE Transactions on Engineering Management 61 (2), 251–260. https://doi.org/10.1109/TEM.2013.2284386.

Hirth, L., 2013. The market value of variable renewables. Energy Economics 38, 218–236. https://doi.org/10.1016/j.eneco.2013.02.004.

Hirth, L., Müller, S., 2016. System-friendly wind power. Energy Economics 56, 51–63. https://doi.org/10.1016/j.eneco.2016.02.016.

Hoffmann, M., Kotzur, L., Stolten, D., Robinius, M., 2020. A Review on Time Series Aggregation Methods for Energy System Models. Energies 13 (3), 641. https://doi.org/10.3390/en13030641.

International Energy Agency, 2018. Electricity information 2018. OECD Publishing, Paris, 690 pp.

Klie, L., Madlener, R., 2020. Optimal Configuration and Diversification of Wind Turbines: A Hybrid Approach to Improve the Penetration of Wind Power. FCN Working Paper. Aachen.

Kotzur, L., Markewitz, P., Robinius, M., Stolten, D., 2018. Impact of different time series aggregation methods on optimal energy system design. Renewable Energy 117, 474–487. https://doi.org/10.1016/j.renene.2017.10.017.

Leipziger Institut für Energie GmbH; Helmut-Schmidt-Universität; BioConsult SH GmbH & Co.KG, 2014. Vorbereitung und Begleitung der Erstellung des Erfahrungsberichts 2014 gemäß § 65 EEG: Vorhaben IIeStromerzeugung aus Windenergie, Hamburg, 352 pp. https://www.bmwi.de/Redaktion/DE/Downloads/XYZ/zwischenbericht-vorhaben-2e.pdf?__blob=publicationFile&v=1 (accessed 14 August 2019).

Milligan, G.W., 1980. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. Psychometrika 45 (3), 325–342. https://doi.org/10.1007/BF02293907.

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

Pfenninger, S., 2017. Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability. Applied Energy 197, 1–13. https://doi.org/10.1016/j.apenergy.2017.03.051.

Pineda, S., Morales, J.M., 2018. Chronological Time-Period Clustering for Optimal Capacity Expansion Planning With Storage. IEEE Trans. Power Syst. 33 (6), 7162–7170. https://doi.org/10.1109/TPWRS.2018.2842093.

Pöstges, A., Weber, C., 2019. Time series aggregation – A new methodological approach using the “peak-load-pricing” model. Utilities Policy 59, 100917. https://doi.org/10.1016/j.jup.2019.05.003.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O.P., Tiwari, A., Er, M.J., Ding, W., Lin, C.-T., 2017. A review of clustering techniques and developments. Neurocomputing 267, 664–681. https://doi.org/10.1016/j.neucom.2017.06.053.

Schütz, T., Schiffer, L., Harb, H., Fuchs, M., Müller, D., 2017. Optimal design of energy conversion units and envelopes for residential building retrofits using a comprehensive MILP model. Applied Energy 185, 1–15. https://doi.org/10.1016/j.apenergy.2016.10.049.

Slootweg, J.G., Kling, W.L., 2003. Aggregated modelling of wind parks in plower system dynamics simulations, in: IEEE Bologna Power Tech Conference Proceedings. IEEE Bologna Power Tech, Bologna, Italy. June 23-26, 2003. IEEE, pp. 626–631.

Sowa, T., Vasconcelos, M., Schnettler, A., Metzger, M., Hammer, A., Reischboek, M., Köberle, R., 2016. Method for the operation planning of virtual power plants considering forecasting errors of distributed energy resources. Electr Eng 98 (4), 347–354. https://doi.org/10.1007/s00202-016-0419-9.

Stadler, M., Groissböck, M., Cardoso, G., Marnay, C., 2014. Optimizing Distributed Energy Resources and building retrofits with the strategic DER-CAModel. Applied Energy 132, 557–567. https://doi.org/10.1016/j.apenergy.2014.07.041.

Teichgraeber, H., Brandt, A.R., 2019. Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. Applied Energy 239, 1283–1293. https://doi.org/10.1016/j.apenergy.2019.02.012.

Zatti, M., Gabba, M., Freschini, M., Rossi, M., Gambarotta, A., Morini, M., Martelli, E., 2019. k-MILP: A novel clustering approach to select typical and extreme days for multi-energy systems design optimization. Energy 181, 1051–1063. https://doi.org/10.1016/j.energy.2019.05.044.

Zhang, N., Kang, C., Xu, Q., Jiang, C., Chen, Z., Liu, J., 2013. Modelling and Simulating the Spatio-Temporal Correlations of Clustered Wind Power Using Copula. Journal of Electrical Engineering and Technology 8 (6), 1615–1625. https://doi.org/10.5370/JEET.2013.8.6.1615.

Appendix A: Analytics for scenarios with nodal pricing

To enable the consideration of scenarios with nodal prices and accordingly the influence of grid congestion on optimal aggregation, an extended version of Eq. (12) is presented in Eq. (18). Subsequently, the individual value components are outlined for scenarios with nodal pricing (cf. Eqs. (19) to (23)).

$$AV_{a,i} = \left(\sum_t \bar{\theta}_{\cdot,a,\cdot} p_{t,\cdot} - c_{\cdot}^{ann} + \sum_t \frac{\bar{\theta}_{\cdot,a,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} (\bar{\theta}_{t,\cdot,\cdot} - \bar{\theta}_{\cdot,\cdot,\cdot}) p_{t,\cdot} + \sum_t \bar{\theta}_{\cdot,a,\cdot} \left(\frac{\theta_{t,a,\cdot}}{\bar{\theta}_{\cdot,a,\cdot}} - \frac{\bar{\theta}_{t,\cdot,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \right) p_{t,a} \right.$$

$$+ \sum_t (\theta_{t,a,i} - \bar{\theta}_{t,a,\cdot}) p_{t,a} - (c_i^{ann} - c_{\cdot}^{ann}) \tag{18}$$

$$\left. + \sum_t \frac{\bar{\theta}_{\cdot,a,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \bar{\theta}_{t,\cdot,\cdot} (p_{t,a} - p_{t,\cdot}) \right) K_{a,i}$$

In case of a nodal scenario, the zonal price p_t of the yield component is set to the average of all nodal prices $p_{t,\cdot}$. A utilization of nodal prices within the yield component would reduce its significance compared to scenarios without nodal pricing and therefore is avoided.

$$VC_{a,i}^{yield} = \sum_t \bar{\theta}_{\cdot,a,\cdot} p_{t,\cdot} - c_{\cdot}^{ann} \tag{19}$$

For the nodal scenario the price within the resource component (cf. Eq. (20)) is also set to $p_{t,\cdot}$.

$$VC_{a,i}^{resource} = \sum_t \frac{\bar{\theta}_{\cdot,a,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} (\bar{\theta}_{t,\cdot,\cdot} - \bar{\theta}_{\cdot,\cdot,\cdot}) p_{t,\cdot} \tag{20}$$

To evaluate the value of a site, not only the local conditions should be considered but the geographic location within the energy system as well. Hence, for the nodal scenario, nodal prices are utilized to calculate the site-specific value component as shown in Eq. (21).

$$VC_{a,i}^{site} = \sum_t \bar{\theta}_{\cdot,a,\cdot} \left(\frac{\theta_{t,a,\cdot}}{\bar{\theta}_{\cdot,a,\cdot}} - \frac{\bar{\theta}_{t,\cdot,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \right) p_{t,a} \tag{21}$$

For the technology component, nodal prices $p_{t,a}$ are utilized as well. Because the formulation in Eq. (22) focusses on the difference of technological aspects e.g. the advantages or drawbacks of high/low wind speed turbines, the location-dependent aspects of technology choices should also include locational, i.e. nodal prices.

$$VC_{a,i}^{technology} = \sum_t (\theta_{t,a,i} - \bar{\theta}_{t,a,\cdot}) p_{t,a} - (c_i^{ann} - c_{\cdot}^{ann}) \tag{22}$$

The grid component presented in Eq. (23) only exists in scenarios with nodal prices and expresses the influence of grid restrictions on the value of an investment choice.

$$VC_{a,i}^{grid} = \sum_{t} \frac{\bar{\theta}_{\cdot,a,\cdot}}{\bar{\theta}_{\cdot,\cdot,\cdot}} \bar{\theta}_{t,\cdot,\cdot} \left(p_{t,a} - p_{t,\cdot}\right) \tag{23}$$

Mainly driven by the difference between the nodal and the average price of all sites, it expresses the change in value related to the site-specific variations in prices. As high renewable infeed and missing transfer capacities lead to low prices, the grid component becomes negative or at least very small for the corresponding locations. This component can help to include the impact of grid restrictions when selecting suitable spatial aggregations and ensures the consideration of grid restrictions in the clustering of investment choices described in section 3.4.

Appendix B: Capacity expansion optimization problem

Objective function: minimization of overall system costs

$$\min_{x,K,LL} C$$

$$C = \sum_{i \in i^C} c_i^{inv} K_i^C + \sum_{i,a} c_{i,a}^{inv} K_{i,a}^{RE} + \sum_{t,i} c_i^{var} x_{t,i}^C + \sum_{t,i,a} c_{i,a}^{var} x_{t,i,a}^{RE} + \sum_t c^{LL} LL_t \tag{1}$$

Balance of supply and demand

$$s.t. \ \ \forall t \ \sum_i x_{t,i}^C + \sum_{i,a} x_{t,i,a}^{RE} + LL_t = D_t + S_t \tag{2}$$

Limitation of conventional generation by installed capacity and availability

$$\forall t, i \ x_{t,i}^C \leq \theta_{t,i,r} K_i^C \tag{3}$$

Limitation of renewable generation per area by installed capacity and availability

$$\forall t, i, a \ \ x_{t,i,a}^{RE} \leq \theta_{t,i,a} K_{i,a}^{RE} \tag{4}$$

Limitation of conventional generation capacities by potential per technology class

$$\forall i \ \ K_i^C \leq K_i^{C,max} \tag{5}$$

Limitation of renewable generation capacities by potential per technology class and area

$$\sum_{i \in i^{tcl}} K_{i,a}^{RE} \leq K_{tcl,a}^{RE,max} \tag{6}$$

Political objective: minimum share of electricity produced from renewables

$$r^{RE} \sum_t D_t \leq \sum_{t,i,a} x_{t,i,a}^{RE} \tag{7}$$

Political objective: limitation of CO2 emissions by conventional production

$$\sum_{t,i\in i^{C},r} \frac{x_{t,i,r}^{C}}{\eta_{i,r}} \varepsilon_{i,r}^{CO_2} \leq R^{CO_2} \tag{8}$$

Post processing (not part of the optimization): calculation of renewable curtailment

$$\forall t, a \;\; S_{t,a} = \sum_{i,a} \left(\theta_{t,i,a} K_{i,a}^{RE} - x_{t,i,a}^{RE} \right) \tag{9}$$

Correspondence

M.Sc. Arne Pöstges

Tel. +49 201 183-6862

E-Mail arne.poestges@uni-due.de

Prof. Dr. Christoph Weber

Tel. +49 201 183-2966

E-Mail christoph.weber@uni-due.de

Chair for Management Science and
Energy Economics
House of Energy Markets and Finance
University of Duisburg-Essen
Campus Essen
Universitätsstr. 2, 45141 Essen, Germany

Tel. +49 201 183-2309

Fax +49 201 183-2703

E-Mail web.hemf@wiwi.uni-due.de

Web www.hemf.net