

# Wasserstein-Distance-Based Temporal Clustering for Capacity-Expansion Planning in Power Systems

Lucas Condeixa

Mathematics and System Analysis  
Aalto University  
Espoo, Finland  
lucas.condeixa@aalto.fi

Fabricio Oliveira

Mathematics and System Analysis  
Aalto University  
Espoo, Finland  
fabricio.oliveira@aalto.fi

Afzal S. Siddiqui

Department of Computer and Systems Sciences  
Stockholm University  
Stockholm, Sweden  
asiddiq@dsv.su.se

**Abstract**—As variable renewable energy sources are steadily incorporated in European power systems, the need for higher temporal resolution in capacity-expansion models also increases.

Naturally, there exists a trade-off between the amount of temporal data used to plan power systems for decades ahead and time resolution needed to represent renewable energy variability accurately. We propose the use of the Wasserstein distance as a measure of cluster discrepancy using it to cluster demand, wind availability, and solar availability data. When compared to the Euclidean distance and the maximal distance, the hierarchical clustering performed using the Wasserstein distance leads to capacity-expansion planning that 1) more accurately estimates system costs and 2) more efficiently adopts storage resources. Numerical results indicate an improvement in cost estimation by up to 5% vis-à-vis the Euclidean distance and a reduction of storage investment that is equivalent to nearly 100% of the installed capacity under the benchmark full time resolution.

## C. Variables

$p_{g,t}$	Dispatch of technology $g$ at time $t$ (MW).
$\bar{b}$	Battery capacity (MW).
$b_t$	State of charge in battery at time $t$ (MW).
$\bar{p}_g$	Installed capacity of technology $g$ (MW).
$\sigma_t$	Load shedding at time $t$ (MW).

## D. Abbreviations

CEP	Capacity-expansion planning
ED	Euclidean distance
FTR	Full time resolution
MD	Maximal distance
TA	Time aggregation
WD	Wasserstein distance

## NOMENCLATURE

### A. Indexes and sets

$g$	Generation technology ( $g \in G$ ).
$t$	Time cluster ( $t \in T$ ).

### B. Parameters

$Av_{g,t}$	Availability of generation technology $g$ at time $t$ ( $[0,1]$ ).
$D_t$	Demand at time $t$ (MW).
$GC_g^G$	Power generation (running) cost of technology $g$ (k€ / MW).
$I_g^G$	Annualized investment needed to install 1 MW of technology $g$ (k€ / MW).
$I^S$	Storage technology annualized investment cost (k€ / MW).
$M_g$	Annual maintenance cost of technology $g$ (k€ / MW).
$RES$	Minimum renewable energy share (accounted for over the entire time horizon) ( $[0,1]$ ).
$S_{lim}$	Percentage limit on load shedding ( $[0,1]$ ).
$S_{effic}$	Battery efficiency for charging and discharging.
$SC^G$	Load-shedding cost (k€ / MW).
$\tau_t$	Length of the time cluster $t$ (hours).

## I. INTRODUCTION

### A. Motivation and Background

The European power sector is undergoing significant and enduring changes. Decarbonization was posed as one of the five European Union headline strategic 2020 targets for smart, sustainable, and inclusive growth. An outcome of this strategy is the target of having 32% of the 28 EU Member States' energy supply provided by renewable sources by 2030 [1], which implies a deep decarbonization of the power sector as heat and transportation sectors increasingly electrify. As a consequence, an increase of almost 100% in installation of renewable capacity took place between 2004 and 2018 [1], of which the major part was in the form of wind and solar photovoltaic power.

Given their dependence on natural phenomena and subsequent intermittency, wind and solar sources are referred to as variable renewable energy sources (VRES). Nowadays, the main barriers to further development of VRES [2] relate to: i) the (generation–transmission–distribution) grid capacity for absorbing intermittent electricity generation; ii) geographical dispersion of sources; and iii) long distances between high VRES potential and high demand locations. Hence, the use of mathematical optimization models that can incorporate variability underpins sound decision making. Geographically extensive power systems – such as the one envisioned by the Energy Union [3] – require similarly expansive and, often overwhelming, amounts of data to be handled. To circumvent

the eventual computational intractability of massive datasets in long-term planning, energy-system models systematically benefit from methodologies to reduce the amount of input data [4], such as temporal aggregation (TA). TA generally consists of grouping periods with similar features (such as demand levels, wind-generation capacity, and so forth) such that decisions are based on time clusters, e.g., representative hours, days, or weeks, that are fewer in number.

Capacity-expansion planning (CEP) involving investment decisions (including for VRES capacity) is affected by operational efficiency as the grid’s structure limits operational flexibility. Thus, joint optimization of investment and operational decisions is desirable. In the presence of high temporal resolution (e.g., hourly data), TA plays an important role, allowing for the consideration of decisions comprising both CEP long-term decisions and (short-term) operational decisions. The need for a high temporal resolution is a consequence of increased VRES use as operational decisions can change in a matter of minutes following VRES availability.

Integrated energy systems’ decisions demand large amounts of data for the implementation of optimization models that can synchronize decisions between regions and over time. Therefore, the importance of performing reliable and accurate TA rests on the fact that the full time resolution (FTR) implementation often cannot be solved in a tractable computational time to support decision making. Thus, any improvement to TA quality can potentially yield benefits for system planning as it allows decision making to consider chronology without compromising computational tractability. Hence, long-term planning that relies on short-term conditions, e.g., VRES availability, requires a TA that represents spatio-temporal features of the system without jeopardizing computational tractability.

## B. Relevant Literature and Contributions

One of the most developed family of methods to perform TA is time-series clustering. These methods are classified as *divisive* when all elements start belonging to the same cluster and new clusters are formed by splitting the original one. Clustering methods can also be classified as *agglomerative* when each element is assigned to its own cluster and neighboring clusters are iteratively merged until a certain clustering level is reached. In each iteration, *hierarchical agglomerative clustering* merges the two most similar, or least discrepant, clusters until a number of clusters initially specified is reached. This set of clusters is then used as a representation of the original time series. In order to group clusters, clustering methods rely on a measure of discrepancy to assess pairwise differences between clusters [4].

In the context of CEP in power systems, employing hierarchical clustering for choosing representative periods is useful for considering time-correlated decisions that rely on time continuity, e.g., storage levels over time or unit-commitment decisions. When performing TA, time clusters can be formed with time periods that are not necessarily sequential. As hierarchical clustering merges neighboring clusters, those time clusters derived from this method can be used in an optimiza-

tion model while maintaining intra-cluster time continuity, cf. [5], [6]. In [5], hierarchical clustering is used with the minimal-maximal distance to assess clusters’ discrepancy. The authors solve a CEP using 30 representative days selected by their clustering method measuring discrepancy as previously proposed by [7]. Similarly, [6] uses hierarchically clustered time periods to assess a CEP also considering storage investments. It models storage decisions to make possible the use of both representative days and representative weeks. As a conclusion, the authors claim that their method obtained solutions closest to the FTR benchmark considering all of the tested clustering methods.

A new discrepancy metric can support the definition of suitable time clusters for CEP optimization, especially when VRES must be included in the system. The Wasserstein distance (WD) is proposed in this paper as a discrepancy measure for the hierarchical clustering method in order to improve the TA performance. The metric can be straightforwardly calculated by solving a transportation problem [8]. Traditionally, WD has been used in a statistical context to compare two sets of probability distribution functions, continuous or discrete. A formal definition as well as its convergence proof can be found in [9]. Among different applications, WD, also referred to as the Kantorovich–Rubinstein distance, has been used to select representative scenarios for stochastic programming [10], within machine-learning applications such as the classification problem [11] and the imaging problem [12], and to support distributionally robust optimization [13].

Our aim is to use WD to improve the quality of clustering of time series typically used for CEP. To assess the benefits of employing the proposed methods, we develop a case study aggregating demand levels as well as wind and solar availability for three distinct single-node systems. By doing so, we illustrate the suitability of the metric in diverse power systems. We assess the performance of the clustering method using different discrepancy measures and compare the results in terms of their correspondence to the decisions made by a benchmark model using FTR. In particular, we find that for CEP problem instances involving both generation and storage adoption with VRES targets, WD outperforms the Euclidean distance (ED) and maximal distance (MD) discrepancy measures in terms of estimating both system cost and installed storage capacity benchmarked to an FTR implementation. Considering previous TA techniques in long-term power-system models, we provide a methodological contribution via a novel clustering discrepancy metric that reinforces the chronology of operational decisions. Our innovation is especially pertinent in the context of VRES penetration in systems with intraday linking constraints such as storage levels.

## II. METHODOLOGY

### A. Hierarchical Clustering using the Wasserstein Distance

The main objective of using a hierarchical clustering method is to form sets of data points that can be represented via a single value or vector of values. When applied to a time series, we obtain a reduced time series with  $k$  time clusters

from an original series of  $n$  time periods. Each cluster is represented by centroids - defined by the mean of the data points belonging to that cluster - or medoids - the data point that is closest to all other data points within the same cluster. The centroids/medoids are used to derive a representation of the original time series with dimensionality reduced from  $n$  to  $k$  values. The hierarchical clustering algorithm consists of the following steps:

- 0) Assume each time period from the FTR is a cluster comprising a single period. The number of clusters then ( $\ell$ ) equals the number of periods in the time series ( $n$ );
- 1) Compute the representative center of each cluster. In case centroids are used, this consists of the mean of the cluster's data points;
- 2) Compute the distance between neighboring clusters (from 1 to  $\ell - 1$ );
- 3) Find the two neighboring clusters with the minimal distance between each other and merge them into one cluster;
  - 3.1) In the case of more than a pair of clusters with the same distance between each other, choose a criterion to select the two to be merged (e.g., the lower-ordered clusters first); and
- 4) If the desired number of clusters is achieved ( $\ell = k$ ), then stop; otherwise, go back to step 1).

In our numerical examples, the number of clusters,  $k$ , is that to be used by the CEP, which we refer to as the clustering level. Therefore, we reduce the original  $n$  time periods to  $k$  time clusters through an iterative routine to update the clusters to which they belong. In practice, a vector called *classes* with dimension  $n$  is created to account for the data points' assignment. Both the vector of classes and the clusters' centroids are the output of this algorithm. The algorithm's novelty comes from the use of the WD to measure the discrepancy between two consecutive clusters. The resulting routine is formally described in Algorithm 1.

In the proposed algorithm,  $n$  represents the length of the FTR time series,  $a_o$  with  $o \in \{1, \dots, m\}$ . The element *class* is the vector with the classes of each data point in  $a_o$  and *k-cent* stores the centroids of each cluster between 1 and  $\ell$ . *dist* stores the distance between clusters  $i$  and  $i + 1$ . The auxiliary matrix,  $\overline{\text{class}}_{[n \times \ell]}$ , stores the values of the classes for each possible aggregation between clusters, and  $\overline{\text{class}}^R$  represents the column of this matrix when cluster  $R$  is aggregated with  $R + 1$ .  $\bar{a}_o^R[j]$  represents how the previous iteration's clustered time series would change in case the  $R^{\text{th}}$  cluster were merged with its immediate neighbor. Then,  $\text{WDO}(\cdot)$  calculates the WD between the original series and the representative centroids in the FTR  $n$  when cluster  $i$  is merged with cluster  $i + 1$ .  $\text{WDO}$  stands for the Wasserstein distance optimization problem being solved to calculate the WD when comparing two time series, in this case  $a_o$  and  $\bar{a}_o^R$ . The intuition is that WD provides a good proxy to measure the representation quality of the clustering to be performed when contrasted with the FTR series. For example, if aggregating time cluster  $i$  with time

---

**Algorithm 1** Hierarchical clustering for  $m$  time series ( $a_1, \dots, a_m$ ) using Wasserstein distance

---

**Input:**  $k, (a_1, \dots, a_m)$

**Output:** classes, centroids

*Initialization :*

- 1: class = (1 : n)
- 2: k-cent<sub>o</sub> =  $a_o$
- 3: **for**  $\ell = n$  to  $k$  **do**
- 4:   **for**  $i = 1$  to  $\ell - 1$  **do**
- 5:     **for**  $j = 1$  to  $n, R = 1$  to  $\ell - 1, o = 1$  to  $m$  **do**
- 6:        $\overline{\text{class}}^R := \text{class}$
- 7:       **for**  $s = R + 1$  to  $n$  **do**
- 8:          $\overline{\text{class}}^R(s) := \text{class}(s) - 1$
- 9:       **end for**
- 10:        $\bar{a}_o^R(j) = \text{mean}(a_o \mid \overline{\text{class}}^R = \overline{\text{class}}^R(j))$
- 11:     **end for**
- 12:     *Computing clusters' discrepancy :*
- 13:      $\text{dist}^{\text{WD}}(i) = \sum_{o=1}^m \text{WDO}(\bar{a}_o^{R=i}, a_o)$
- 14:     *Electing new classes :*
- 15:      $\text{min} = \text{argmin}_{x \in 1:\ell}(\text{dist}^{\text{WD}}(x))$
- 16:     **for**  $j = \text{min} + 1$  to  $n$  **do**
- 17:       class( $j$ ) = class( $j$ ) - 1
- 18:     **end for**
- 19:     *Updating centroids :*
- 20:     **for**  $i = 1$  to  $\ell, o = 1$  to  $m$  **do**
- 21:       k-cent<sub>o</sub>( $i$ ) =  $\text{mean}(a_o \mid \text{class} = \text{class}(i))$
- 22:     **end for**
- 23: **end for**
- 24: **return** class, k-cent<sub>o</sub>

---

cluster  $i + 1$  yields a larger WD to the original series than when aggregating  $i^*$  with cluster  $i^* + 1$ , then it is preferable to aggregate  $i^*$  instead of  $i$ . By using the WD, we expect to retain, as much as possible, the time series' information while reducing the temporal resolution by grouping  $i^*$  and  $i^* + 1$ . In lines 10 and 19, we use a conditional mean -  $\text{mean}(\cdot \mid \cdot)$  - to calculate the centroids belonging to a specific cluster, thereby averaging just the data points belonging to the same class. The algorithm terminates when the number of clusters,  $k$ , is achieved. The methodology and, therefore, the algorithm can be trivially generalized to an  $m$  number of series - in our examples, we used 3 series for 3 nodes, but more temporal features of the system and/or nodes can be used - thus making it applicable also to larger test systems.

Line 12 of Algorithm 1 computes the discrepancy between clusters. Other possible cluster discrepancy metrics are ED and MD. ED is the norm of the difference between corresponding centroids of each cluster and its following neighbor:

$$\text{dist}^{\text{ED}}(i) = \sum_{o=1}^m \|\text{k-cent}_o(i) - \text{k-cent}_o(i + 1)\|$$

The second metric considers the maximal difference between two elements belonging to neighboring clusters. It then considers the two points farthest away in the case a cluster  $i$  is clustered with its immediate neighbor  $i + 1$ . The MD is

then defined as:

$$\text{dist}^{\text{MD}}(i) = \sum_{o=1}^m \max \|a_o^i - a_o^{i+1}\|$$

Examples of the employment of hierarchical clustering using ED and MD can be found, respectively, in [5] and [6]. In Section IV, we perform the clustering using the three metrics presented and compare the results obtained from the CEP model described next.

### B. CEP Optimization Model

Consider a CEP model that decides the infrastructure to be installed to fulfil the demand of a given region. The geographical scope of the model can vary from an independent distribution system to a multi-country integrated network. For the sake of simplicity, we consider a single-node system disregarding electricity transmission. We also consider a greenfield standpoint, meaning that no capacity is initially available. Storage capacity is also considered with only battery storage available. Additionally, to meet sustainability requirements of future power systems, a minimum renewable energy penetration is imposed on the system, calculated as a percentage of the total production within the planning horizon. The CEP problem is formulated as follows:

$$\min_{\Xi} \sum_g (I_g^G + M_g) \bar{p}_g + \sum_{g,t} GC_g^G p_{g,t} \tau_t + \sum_t SC^G \sigma_t \tau_t + I^S \bar{b} \quad (1)$$

$$\text{s.t.} \quad \sum_g p_{g,t} + \sigma_t - S_{\text{effic}}(b_t - b_{t-1}) = D_t, \quad \forall t \in T \quad (2)$$

$$p_{g,t} \leq Av_{g,t} \bar{p}_g, \quad \forall g \in G, t \in T \quad (3)$$

$$\sum_{g^* \in G^R, t} p_{g^*,t} \geq RES \sum_{g \in G, t} p_{g,t} \quad (4)$$

$$\sigma_t \leq S_{\text{lim}} D_t, \quad \forall t \in T \quad (5)$$

$$b_t \leq \bar{b}, \quad \forall t \in T \quad (6)$$

$$b_1 = b_{T(\text{end})} \quad (7)$$

$$p_{g,t} \geq 0, \forall g \in G, t \in T; \bar{p}_g \geq 0, \forall g \in G; \sigma_t, b_t \geq 0, \forall t \in T; \bar{b} \geq 0, \quad (8)$$

where  $\Xi \equiv \{p_{g,t}, \bar{p}_g, \sigma_t, b_t, \bar{b}\}$ . The objective function in (1) minimizes generation-investment costs (accounting for both the annualized cost of investment and the maintenance costs) plus the operational costs of generation, load-shedding costs, and battery-investment costs. Constraints (2) define the energy balance of the system. Constraints (3) limit the generation to the available capacity (while considering that non-renewable generators always have 100% availability). Constraints (4) enforce the production of a minimal share of renewable energy. Constraints (5) limit the load shedding to a certain percentage of the demand. Constraints (6) limit the amount of energy stored according to the storage capacity. Constraints (7)

enforce that the storage level at the end of the time horizon is the same as the initial level. Finally, (8) sets the domain of the decision variables.

The model presented considers joint investment and operational decisions in which some features were on purpose disregarded, e.g., unit commitment, ramping limits, voltage angles, etc. By using a less-detailed operational description of the problem, we aim to focus on the effects of temporal reduction, i.e., we reduce the operational complexity to focus on the long-term investment decisions. An underlying operational assumption is that all generation plants and storage are fully flexible, viz., they can vary their production freely between 0 and their capacity.

Nonetheless, using a simple CEP optimization model, we can still assess important features related to the clustering such as how the total costs are estimated when using a clustered approach and how capacity-investment decisions in power systems can differ among clustering approaches.

### III. CASE STUDY

We model three different single-node systems in order to contrast the decisions made when considering different nodal profiles in terms of demand and wind and solar availability. We consider three different time series for each node, viz., demand, wind availability, and solar availability. Node 1 ( $N1$ ) has the highest demand but a relatively low wind and solar availability on average. Node 2 ( $N2$ ) has a high wind availability, whereas Node 3 ( $N3$ ) has the highest solar incidence of them all (see Figure 1).

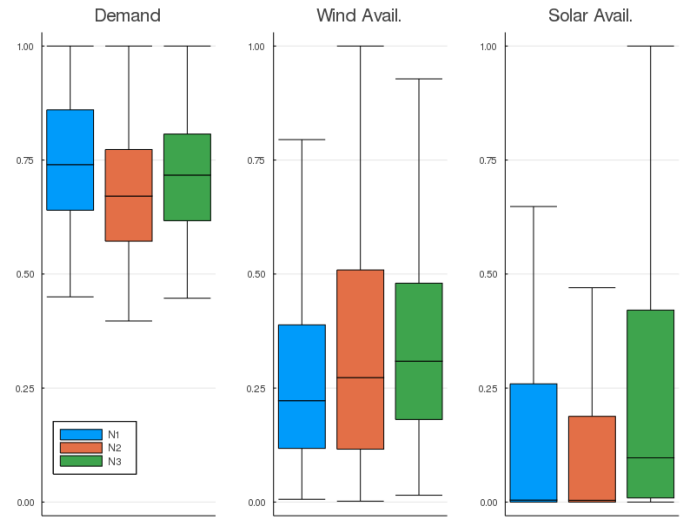


Fig. 1: Demand, wind availability, and solar availability at nodes  $N1$  (blue),  $N2$  (orange), and  $N3$  (green)

Operational details intrinsic to power system such as which distribution lines to use or the reserve systems needed to guarantee the system's stability are outside the scope of this study. As the nodes are supposed to share the same geographical region, the time series are influenced by some common seasonal weather aspects.

The study encompasses nine time series (three nodes with three features), comprising a full year with hourly resolution, i.e., 8,760 hours. Each time series ranges from 0 to 1, as they are normalized. Every clustering performed aims to replace the original 8,760 hours FTR by a set of representative hours per month to reduce the model size. The series are analyzed on a monthly basis to maintain part of the seasonal effects throughout the year. For example, if we set the number of clusters  $k$  to 2, we use 2 representative hours per month, implying 24 representative hours for the entire year. Our experiments range from 1 cluster per month, with 12 representative hours, to 100, comprising 1,200 representative hours.

The data used for demand, wind availability, and solar availability were collected from ENTSO-E databases of Denmark, Germany, and Spain [14]. No public economic incentives or premiums (such as the ones to promote renewable capacity expansion) are included in the model. Except from the minimal renewable generation share - which in the case study is 50%. We assume perfect competition, i.e., producers have no market power to manipulate prices. Five generation types can be employed at each node: wind, solar, coal, open-cycle gas turbines (OCGT), and closed-cycle gas turbines (CCGT). Battery storage investments are also possible in the system.

#### IV. RESULTS AND DISCUSSION

Figure 2 shows the relative difference between the FTR total costs and the costs found when time clusters are considered. The relative difference is computed by the difference between total costs found by the clustered optimization and the FTR total costs divided by the latter. The total cost serves as an indicator for investors and decision makers about how efficiently resources are being allocated. Errors in the total cost, both under- and overestimation, lead to either the increase of real costs to cover future corrections needed because of insufficient capacity or to financial waste due to underused structures. Therefore, as long as all of the constraints are satisfied, cost estimation reflects the decisions' accuracy.

As can be observed, MD performs the worst among the three measures. One reason for that relates to the representation of the maximum difference between elements within a cluster. Whereas the average can provide descriptive statistics about the data points, the difference between the highest and the lowest levels fails to convey information about statistical moments. In general, MD clustering leads to a less conservative approach as the extreme cases, in terms of high demand and low wind and solar availability, can only be accurately represented when a large number of clusters is used. These cases are averaged with more favorable hours, i.e., low demand and high wind and solar availability, leading to lower capacity investments than required by the system. Therefore, considering clustering levels ranging from 1 to 100 clusters per month, the total costs resulting from MD clustering are considerably underestimated.

At Node 2, both ED and WD lead to similar costs along the full range of clustering levels. Nevertheless, at Nodes 1 and 3, WD performs consistently better than ED. When the

clustering level reaches 25 clusters per month, the difference between using ED and WD increases to 5.2%.

One of the reasons for that is a relatively high solar availability at both Nodes 1 and 3 and how the representative values of the cluster are set. For example, when morning or afternoon hours are clustered with night hours, this set of hours will be represented by the average. In contrast with night hours in the FTR, this new cluster created can present solar availability higher than zero. Thus, the optimal clustered solution can assign solar generation for those hours, whereas storage or other generation sources will be demanded in the FTR as the solar availability is concentrated within daylight hours. We could notice that solar seasonal availability seems to be less perceived by the clustering procedure under ED than when WD is used.

Differences between ED and WD as discrepancy metrics can be seen also when analyzing the storage investment (see Figure 3). We focus on the effects of using only WD and ED since employing MD as a discrepancy metric was considerably less effective in terms of total costs. At all nodes, the clustered time periods lead to an overestimation of the storage capacity needed. For Nodes 1 and 3, following their solar availability, WD leads to more accurate storage investments, mainly for clustering levels below 25, i.e., around one representative day per month. Savings at Node 1 when using WD vis-à-vis ED for hierarchical clustering can be up to 25,000 MW of battery capacity, which is about 100% of the FTR battery capacity. These benefits arising from using WD are a consequence of a better assessment of the storage needed to enable further wind and solar penetration.

#### V. CONCLUSIONS

In this study, we propose the Wasserstein distance as a discrepancy metric for hierarchical clustering. We use the metric as a discrepancy measure for the clustering process taking advantage of its chronological properties. We conclude that in some renewable energy profiles, such as with a high incidence of insolation, CEP optimization can benefit from the use of WD to select clusters effectively.

While using clusters screened via WD, the optimized decision improves the cost estimation in up to 5.2% of the FTR system's full cost when compared to the results if clusters formed using the ED are used. Storage capacity estimation also benefits from the use of WD by increasing accuracy vis-à-vis ED up to 100% of the FTR solution. Despite this success, not all generation technologies could benefit equally from the use of WD to improve the clustering resolution. Future work to explore how statistical moments can affect WD's accuracy would be warranted in order to reveal conditions in which WD can capture better the features to support decision making in CEP optimization.

Different clustering methods can be used to improve decisions such as the storage capacity. As indicated in [15], Enhanced Representative Periods (ERP) can lead to better energy storage systems planning. Therefore, the effects of using WD for performing clustering in such methods and

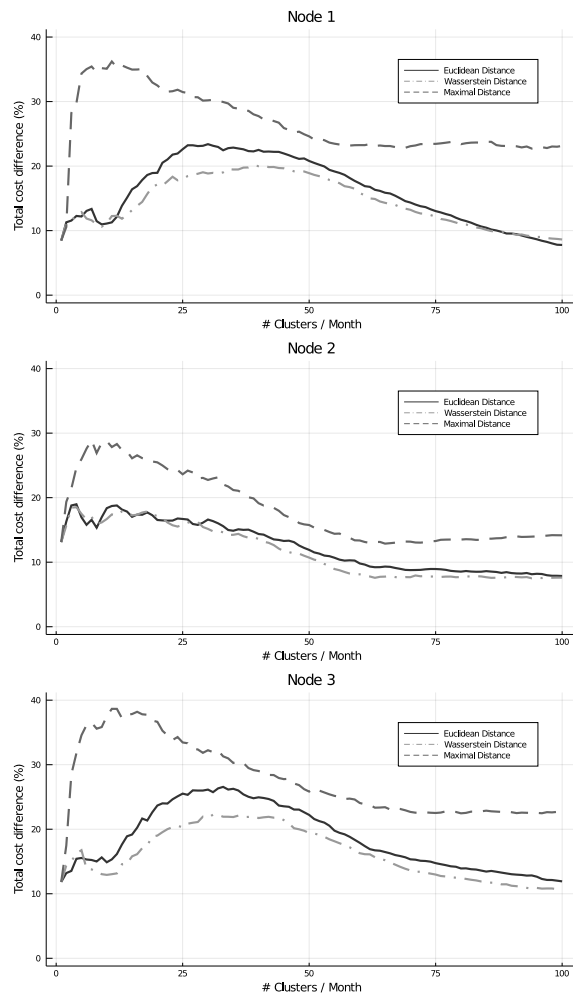


Fig. 2: Percentage difference between the costs as estimated by clustered time-series and the FTR cost, per node

the respective outcomes on energy storage planning are also directions for future work. Additionally, an extension of the numerical examples to an integrated multi-nodal system can inform how WD relates to the temporal correlation between different regions with complementary weather conditions, such as the effects of having a region with a high solar incidence connected to another region with high wind availability. Finally, to provide further evidence of the WD performance as a discrepancy measure, analysis with alternative measures could also be undertaken.

## REFERENCES

- [1] Eurostat, "Renewable energy statistics," ec.europa.eu, Mar. 6, 2020. [Online] Available: [www.ec.europa.eu/eurostat](http://www.ec.europa.eu/eurostat)[Accessed: April 16, 2020].
- [2] European Court of Auditors, "Wind and Solar Power for Electricity Eeneration: Significant Action Needed if EU Targets to be Met," Special report, no. 08, pp. 1–58, 2019. Available: [www.eca.europa.eu](http://www.eca.europa.eu)[Accessed: April 16, 2020].
- [3] European Commission, "Fourth Report on the State of the Energy Union," EC-COM(2019) 175, 2019. Available: [www.ec.europa.eu](http://www.ec.europa.eu)[Accessed: April 16, 2020].
- [4] H. Teichgraber and A. R. Brandt, "Clustering methods to find representative periods for the optimization of energy systems: An initial

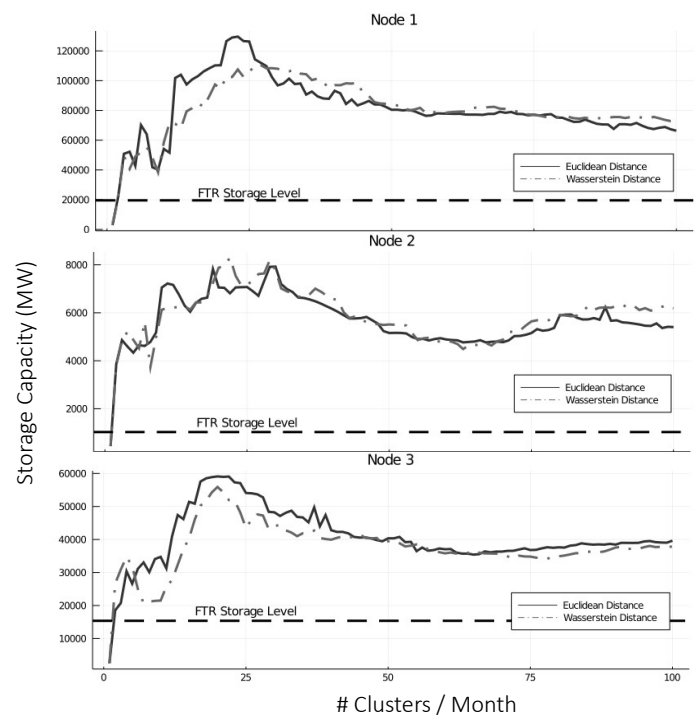


Fig. 3: Storage capacity at each node and the respective FTR solution

- framework and comparison," *Applied energy*, vol. 239, pp. 1283–1293, 2019.
- [5] Y. Liu, R. Sioshansi, and A. J. Conejo, "Hierarchical Clustering to Find Representative Operating Periods for Capacity-Expansion Modeling," *IEEE Transactions on Power Systems*, vol. 33, 3, pp. 3029–3039, 2018.
- [6] S. Pineda and J. M. Morales, "Chronological Time-Period Clustering for Optimal Capacity Expansion Planning with Storage," *IEEE Transactions on Power Systems*, vol. 33, 6, pp. 7162–7170, 2018.
- [7] J. H. Ward Jr., "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, vol. 58, 301, pp. 236–244, 1963.
- [8] W. Römisch, "Stability of Stochastic Programming Problems," in Ruszczyński, A. and Shapiro, A. (eds.), *Stochastic Programming, Handbooks in Operations Research and Management Science*, vol. 10, pp. 483–554, 2003.
- [9] X. Nguyen, "Convergence of Latent Mixing Measures in Finite and Infinite Mixture Models," *The Annals of Statistics*, vol. 41, 1, pp. 370–400, 2013.
- [10] H. Heitsch and W Römisch, "Scenario Reduction Algorithms in Stochastic Programming," *Computational Optimization and Applications*, vol. 24, pp. 187–206, 2003.
- [11] L. Oudre, J. Jakubowicz, P. Bianchi, and C. Simon, "Classification of Periodic Activities Using the Wasserstein Distance," *IEEE Transactions on Biomedical Engineering*, vol. 59, 6, pp. 1610–1619, 2012.
- [12] J. Lellmann, D. A. Lorenz, C. Schonlieb, and T. Valkonen, "Imaging with Kantorovich–Rubinstein Discrepancy," *SIAM Journal of Imaging Sciences*, vol. 7, 4, pp. 2833–2859, 2014.
- [13] P. M. Esfahani and D. Kuhn, "Data-Driven Distributionally Robust Optimization using the Wasserstein Metric: Performance Guarantees and Tractable Reformulations," *Mathematical Programming Series A*, vol. 171, pp. 115–166, 2018.
- [14] ENTSO-E data portal. [Online] Available: [www.entsoe.eu/data/data-portal/](http://www.entsoe.eu/data/data-portal/)[Accessed: September 1, 2019].
- [15] S. Wogrin, D. A. Tejada-Arango, S. Pineda and J. M. Morales, "What time-period aggregation method works best for power system operation models with renewables and storage?," 2019 International Conference on Smart Energy Systems and Technologies (SEST), Porto, Portugal, pp. 1–6, 2019.