Novel node selection approach for continent-wide power system studies using spatio-temporal clustering

Maria Krutova
University of Oldenburg,
Oldenburg, Germany
Technical University of Denmark,
Lyngby, Denmark
Email: maria.krutova86@gmail.com
Telephone: +7 (905) 582–3306

Lueder von Bremen
ForWind – Zentrum fuer Windenergieforschung
University of Oldenburg,
Oldenburg, Germany
Email: lueder.von.bremen@forwind.de
Telephone: +49(0)441 798–5071
Fax: +49(0)441 798–5099

Spyros Chatzivasileiadis
Technical University of Denmark
Lyngby, Denmark
Email: spchatz@elektro.dtu.dk
Telephone: +45 53 83 96 58

***Abstract*—Plans to decrease the use of conventional power sources and deploy renewable energies call ideas for global power networks. In such networks the energy would be transferred several by thousand kilometers from regions with excess renewable power generation to ones that cannot fully cover their demand. Continental-wide grid simulation for system studies encounters the problem to allocate representative nodes. In the simplest case the power network is defined as one node per country. This approach is sufficient in Europe due to moderate country sizes, but is too coarse for very large countries (eg. China, Russia, India). In particular in those countries the renewable power potential, population density and therefore demand vary in the spatial scale stronger than in Europe. Additionally, daily peak loads are shifted in countries spanning through several time zones.**

In this work a novel clustering approach is developed to detect regions with similar spatio-temporal characteristics that are assigned nodes for a power system study. A combination of ST-DBSCAN based clustering method and merging algorithm produces sets of clusters that are chosen for DC power flow simulations connecting Europe and Asia.

20 years MERRA wind speed and solar radiance data is used to compute onshore wind power and photovoltaic power. The demand is estimated on population density. The sensitivity and flexibility of the clustering approach are discussed, examples of resulting grids are shown and benefits with respect to reduced balancing needs are shown. It is found that grids derived from renewable power and residual load clusters have the same total backup needs, while the transmission capacity is affected by the size of the grid.

I. Introduction

Studying large scale power systems, where locations of the nodes cannot be assigned to specific generation or load centers, can pose a challenge of calculating the power grid. Assigning a node to a country or a group of nearby countries [1] does not take into account climate and population density differences. The differences are not as evident in Europe due to moderate country sizes. On the contrary large natural areas in Asia (Siberia, Tibet plateau, Taklamakan desert etc.) have sparse population, but still possess high solar and/or wind potential. Using the clustering approach to group similar points allows to detect regions with high renewable potential or demand more precisely.

TABLE I
Power mix

Condition	β_i
Annual solar surface radiation	
less than 1000 kWh/m^2	0.0
1000-1500 kWh/m^2	0.15
1500-2000 kWh/m^2	0.25
more than 2000 kWh/m^2	0.35
Average wind power (equatorial area correction)	
less than 10% of G^{rated}	$1 - 5\frac{\langle G_i^{W(t)}\rangle}{G^{rated}}$

II. Clustering

A. Data

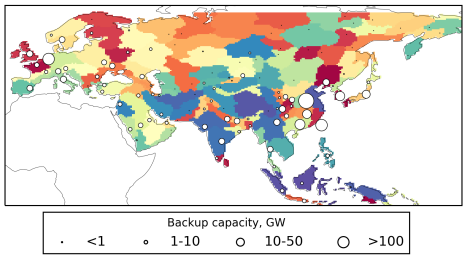
For simplicity only wind and photovoltaic (PV) power are considered in renewable generation. Assuming no significant climate change in the future, feed-in time series are derived from MERRA 1996-2015 wind speed and solar surface radiation [2], [3]. Wind power time series $G_i^W(t)$ are calculated with Enercon E-126 power curve [5]. PV power time series $G_i^{PV}(t)$ are calculated with typical characteristics of PV cells [6], inclination angle is not accounted. Wind power capacities are assumed to be installed uniformly. PV power capacities are scaled to the wind power ones so that

$$\beta_i = \frac{\langle G_i^{PV}(t)\rangle}{\langle G_i(t)\rangle} = \frac{\langle G_i^{PV}(t)\rangle}{\langle G_i^{PV}(t)\rangle + \langle G_i^{W}(t)\rangle} \qquad (1)$$

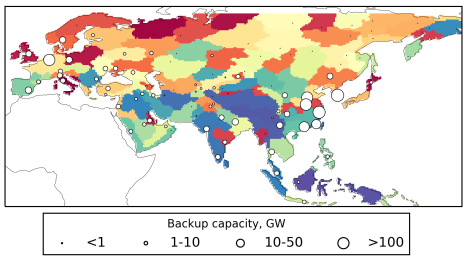
where β_i is chosen depending on annual solar surface radiation (Table I). Additional correction (*) is introduced for South-East Asia, as high solar potential coincides with low equatorial winds, therefore scaling PV to wind power would produce unrealistic results. PV share β takes values only within $[0, 1]$ range.

The difficulties of installing and maintaining wind and PV farms in Tibet and Siberia were not considered, therefore the only generation restriction accounted for those regions is PV power share (1).

Synthetic load time series $L_i(t)$ are calculated based on the average load in MERRA grid cell retaining daily and yearly cycles and accounting increased consumption



(a) Residual load clusters, 115 nodes



(b) Renewable power clusters, 100 nodes

Fig. 1. Isolated nodes backup capacity comparison, $\alpha = 1.0$

during hot/cold weather (reflects high coolers/heaters usage respectively). The average loads per grid cell are derived from population density [9] and country load per capita data [10].

B. Clustering algorithms

Clustering approach is used to detect regions with similar renewable power generation $G_i(t)$ or residual load $R_i(t) = L_i(t) - G_i(t)$. A spatio-temporal algorithm CorClustST inspired by ST-DBSCAN approach [7], [8] handles clustering of the time series. The spatio-temporal neighborhood is defined for distance $R_{max} = 250$ km and time series correlation $\rho_{min} = 0.7$. Originally the algorithm also sets minimum number of points in the neighbourhood to allow noise detection. In this work the limit is disregarded and therefore all points are clustered, no noise is detected. This allows further comparison of the grids derived from renewable power and residual load clusters (Fig. 1.

The total number of detected spatio-temporal clusters may exceed 1,000 clusters for $R_{max} = 250$ km, $\rho_{min} = 0.7$. Converting them to power grid nodes with subsequent grid calculation is a computational heavy problem. The number can be reduced by merging adjacent clusters. The correlation of aggregated time series of the clusters is used as the similarity criteria. The merging is stopped, when total cluster number is reduced to desired limit or the correlation between any adjacent cluster pair is below preset threshold – $\rho^M_{min} = 0.5$. Then each of the resulting clusters is considered as a power grid node. The node location is determined at the residual load center – the subregion center, where the average 20-year residual load is the highest in cluster. Nodes locations do not always match for renewable power and residual load clusters, because of different cluster shape.

III. Power grid calculation

A. Methodology

Each node n is characterized by aggregated generation $G_n(t)$ and load $L_n(t)$ time series. The generation time series are scaled to load to achieve pre-defined global renewables share $\alpha = \langle G \rangle / \langle L \rangle$. Unlike PV share β, global renewables share can take values higher than 1, as generation and load time series are independent.

It is assumed that the node covers its own demand first. Then the excess power is transmitted to nodes with insufficient generation. After the power exchange the mismatch at each node is given as

$$M_n(t) = G_n(t) - L_n(t) + \sum_l f_l(t) \tag{2}$$

The remaining demand ($M_n(t) < 0$) is covered with backup power

$$B_n(t) = -\min(0, M_n(t)) \tag{3}$$

The power flows f_l through links l are calculated in such way that total backup $\sum_n B_n(t)$ and flow dissipation $\sum_l f_l^2$ are minimized at each time step t. This is a graph optimization problem, which was solved using Gurobi with Python interface. [11] The transmission capacity of each link is unlimited.

The backup power accounts for power dispatched from possible storages or other renewable sources. The power from non-renewable sources is calculated for $\alpha < 1.0$ as a constant share C^{NR} of the instant backup required by the node. The dispatch is performed after the power exchange and does not affect the power transmission.

$$B'_n(t) = B_n(t)(1 - C^{NR}) \tag{4}$$

The constant C^{NR} is calculated from global average load to global average backup ratio.

$$C^{NR} = (1 - \alpha)\frac{\langle L \rangle}{\langle B \rangle} \tag{5}$$

For $\alpha \geq 1.0 : C^{NR} = 0, B'_n(t) = B_n(t)$.

Additional quantities are derived from calculated backup and power flow time series. The backup capacity C^B_n is the amount of backup power, that would be sufficient 99% of the time to cover the demand of the node n.

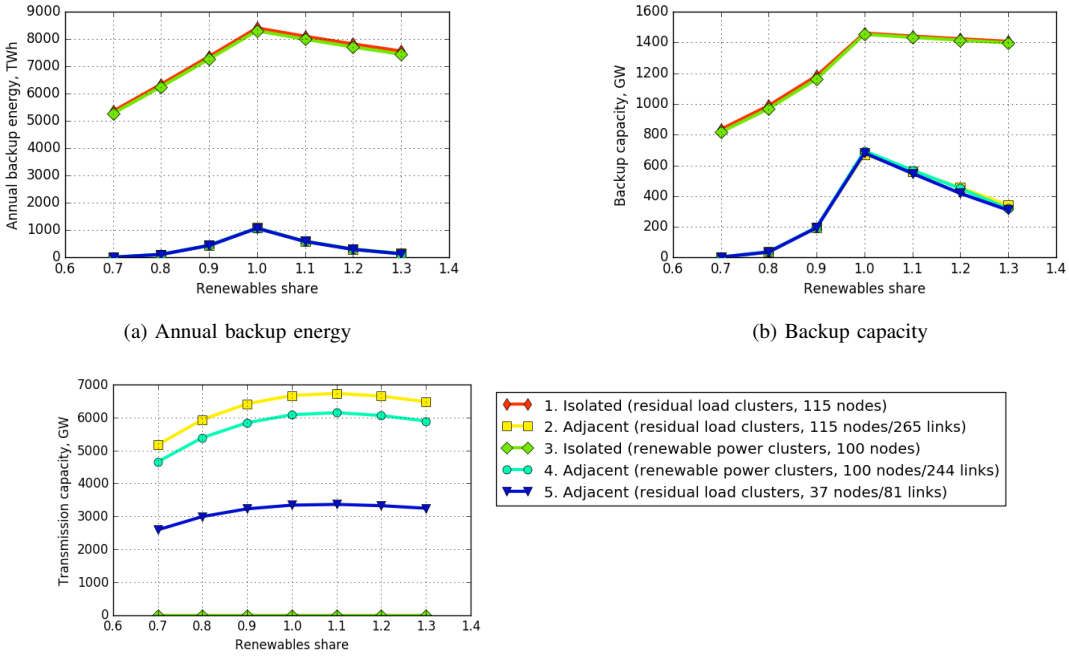
$$0.99 = \int_0^{C^B_n} p(B'_n) dB'_n \tag{6}$$

Similarly the transmission capacity κ^T_l is the required capacity of the link l, that is sufficient to transmit power 99% of the time.

$$0.99 = \int_0^{\kappa^T_l} p(f_l) df_l \tag{7}$$

The link usage is defined as average power transmitted through the link to the transmission capacity of the link l ratio. It reflects, how much of the required link capacity will be used on the average.

$$\Lambda_l = \frac{\langle f_l(t) \rangle}{\kappa^T_l} \cdot 100\% \tag{8}$$



(a) Annual backup energy

(b) Backup capacity

(c) Transmission capacity

Fig. 2. Grid results comparison depending on the grid size and global renewables share α

B. Grid topology

The grid topologies are introduced for three sets of nodes: 115 and 37 nodes derived from residual load clusters, and 100 nodes from renewable power clusters. The results from different grids can be compared, since all points were clustered, i.e. the same data is used to calculate aggregated time series in the nodes.

The base grid topology is constructed by linking nodes that belong to adjacent clusters. The adjacency is defined by a common land border or separation by the sea, if the sea distance is comparable to clusters' sizes. This grid is further referred as 'Adjacent'. The case, where no links are present, i.e. nodes are isolated from each other, is referred as 'Isolated'.

IV. Results

A. Grid size comparison

Regardless of the data clustered, large load centers are detected in West Europe and East Asia for global renewables share $\alpha = \langle G \rangle / \langle L \rangle = 1.0$ (Fig. 1). Additionally nodes requiring noticeable amount of backup power are found in India and Middle East. The backup capacity in the node differs depending whether the power grid was derived from renewable power or residual load clusters. In case of the latter high backup needs are more emphasized, as areas with low/high demand areas are assigned to different clusters (Fig. 1a). Renewable power clustering distributes demand and power generation more equally resulting into nodes with less explicit backup capacity differences (Fig. 1b).

Computational results show, that when the nodes are connected into a grid, the total yearly backup energy needs are strongly reduced in comparison to the isolated nodes case (Fig. 2a). The reduction depends on the global renewables share $\alpha = \langle G \rangle / \langle L \rangle$, but is not significantly affected by the number of nodes in the grid. Additionally part of the demand is covered from non-renewable sources for $\alpha < 1.0$. With the unlimited transmission capacity the annual backup energy reduction is 87.5% and the total backup capacity is reduced by 50% for $\alpha = 1.0$. Total annual demand in the grid can be effectively covered at renewables share $\alpha = 0.7$. Similar backup reduction is observed for increased share value $\alpha = 1.3$ (Fig. 2a), while the backup capacity decreases on a slower rate (Fig. 2b).

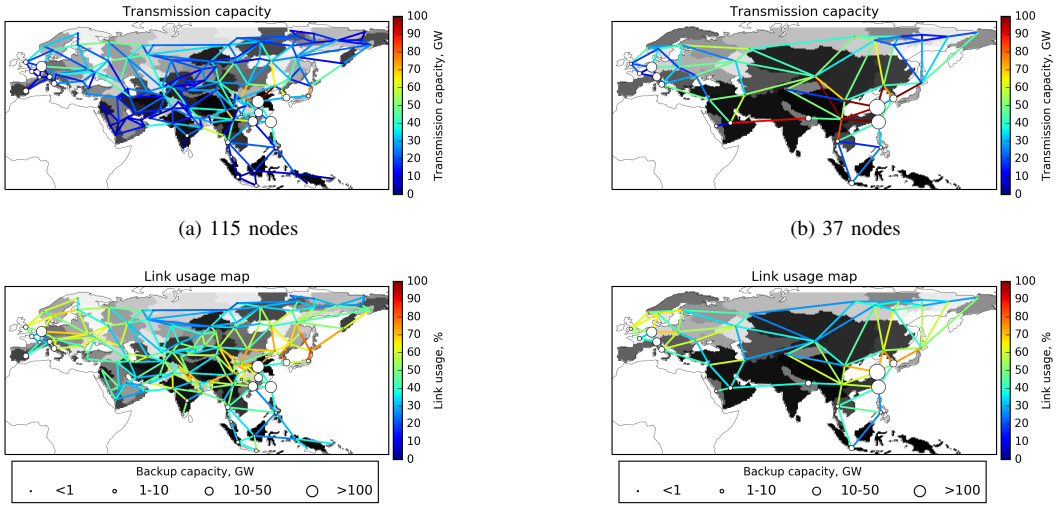
Unlike the backup, the transmission depends on the grid topology. Total transmission capacity decreases with the number of links, though the reduction is not linear. Between 115 and 37 nodes grids the number of links is reduced by 30%, while the transmission capacity reduction is 50% for $\alpha = 1.0$.

Reducing number of nodes(links) leads to increased transmission through the remaining links. 115 nodes/265 links grid resolved for $\alpha = 1.0$ contains a number of links with transmission capacity within 0-30 GW, only a few links located in East China require transmission capacity 80-100 GW (Fig. 3a). Smaller 37 nodes/81 links grid forces several links to transmit more than 100 GW (Fig. 3b).

On the contrary link usage is subjected to less significant changes. The usage map for 115 nodes grid (Fig. 3c) follows the same pattern as the one for a smaller 37 nodes grid (Fig. 3d)) with high usage links located in European and East Asian parts (close to the high load nodes).

B. Renewables share comparison

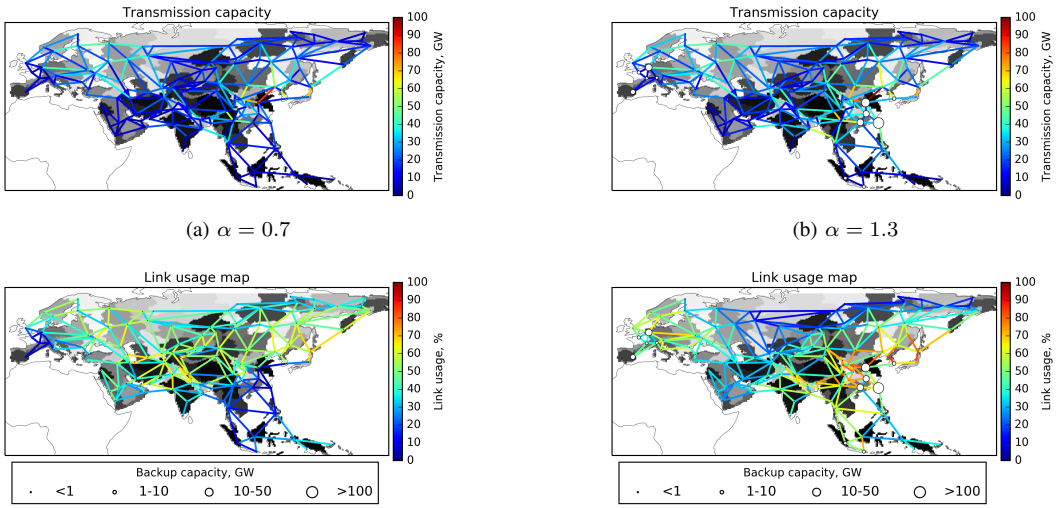
The chosen transmission scheme redistributes only excess renewable power in the grid. The required transmission capacity changes almost equally for all links with the renewables share (Fig. 4a, 4b). On the contrary link usage is sensitive to the amount of renewable power in the grid.



(a) 115 nodes (b) 37 nodes

(c) 115 nodes (d) 37 nodes

Fig. 3. Grid simulation results, grid size comparison for residual load clusters grids: 115 and 37 nodes, $\alpha = 1.0$



(a) $\alpha = 0.7$ (b) $\alpha = 1.3$

(c) $\alpha = 0.7$ (d) $\alpha = 1.3$

Fig. 4. Grid simulation results, global renewables share comparison for residual load clusters grid, 115 nodes

When renewables share is low ($\alpha = 0.7$) the power exchange of the main grid with Southeast Asia and Central Europe nodes becomes less intensive (Fig. 4c). The nature of the decreased usage may be different as European nodes have relatively high load, while Southeast Asian nodes in this model are peculiar for the focus on PV power production due to power mix definition (Table I). The exchange resumes as renewables share increases. With higher renewables share ($\alpha = 1.3$) Far North nodes are less important to the main grid – power exchange with them almost ceases. The nodes are located in the areas with low solar radiation, therefore they provide only wind power (Table I). Additional investigation is needed to determine, whether node production focus is the main factor affecting the intensity of link usage, or node position in the grid and demand also play a role.

V. Conclusion

Combined clustering approaches allow to split large territories into desired number of clusters and use them as nodes in power grid simulations. The number of nodes and links does not affect total backup and backup capacity. Furthermore link usage and node backup capacity distribution are almost invariant with grid topology.

Link transmission capacity results, however, depend on the grid. Overall less power is transmitted between nodes in the smaller grid (Fig. 2c), while the required transmission capacity is significantly increased for the selected links (Fig. 3a, 3b). The link usage appears to be particularly sensitive to the amount of renewable power in the grid. The results also do not take into account the transmission processes going inside the cluster and are a subject of future research.

Acknowledgment

This work is based on the graduation thesis of the European Wind Energy Master program. The work presented in this study is supported by the Ministry of Science and

Culture of Lower Saxony within the project ventus efficiens (ZN3024, MWK Hannover).

REFERENCES

[1] M. Krutova, A. Kies, B. Schyska, and L. von Bremen, *The Smoothing Effect for Renewable Resources in an Afro-Eurasian Power Grid*, Advances in Science and Research, 2017

[2] Global Modeling and Assimilation Office (GMAO) (2008), *tavg1_2d_slv_Nx: MERRA 2D IAU Diagnostic, Single Level Meteorology, Time Average 1-hourly V5.2.0*, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed [07/01/2017] 10.5067/B6DQZQLSFDLH

[3] Global Modeling and Assimilation Office (GMAO) (2008), *tavg1_2d_rad_Nx: MERRA 2D IAU Diagnostic, Radiation Surface and TOA, Time Average 1-hourly V5.2.0*, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed [07/01/2017] 10.5067/RI9VTUQN74XJ

[4] K. Kawajiri, T. Oozeki and Y. Genchi. *Effect of Temperature on PV Potential in the World*. Environmental Science & Technology 45.20, 2011

[5] ENERCON, *ENERCON product overview*, 2016.

[6] E. Skoplaki and J.A. Palyvos. *On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations*. Solar Energy 83.5, pp. 614-624, 2009

[7] D. Birant and A. Kut, *ST-DBSCAN: An algorithm for clustering spatial-temporal data*. Data & Knowledge Engineering 60 (1), 2007

[8] M. Huesch, *CorClustST – Correlation-based Clustering of Big Spatio-Temporal Datasets*, Future Generation Computer Systems, 2017 (paper under review)

[9] Center for International Earth Science Information Network - CIESIN - Columbia University. *Gridded Population of the World, Version 4 (GPWv4): Population Density*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC) Accessed [09/03/2017] 10.7927/H4HX19NJ

[10] IEA Policies and Measures Database. *Key World Energy Statistics 2016*, 2016

[11] Gurobi Optimization Inc. *Gurobi Optimizer Reference Manual*. 2015