Method Article

Using the max-p regions problem algorithm to define regions for energy system modelling

Christian Etienne Fleischer
Europa-Universität Flensburg Germany

ABSTRACT

The input data of energy system models are, in many cases, aggregated data used to represent regions of interest. This paper presents a method of using energy-related spatial data and the max-p-regions method to define regions. The method aims to assign areas to regions that are similar in how much they consume, produce and store electricity.

- A spatial dataset of administrative areas for 30 European countries is presented and used to apply the proposed method.
- Use of energy-related spatial data to define regions of energy system model.
- The method uses the heuristic solution of the max-p regions problem to reduce the spatial resolution of energy system models.

© 2021 The Author. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

ARTICLE INFO
Method name: Defining regions according to energy-related attributes using the max-p-regions method for energy system modelling
Keywords: Max-p-regions, Spatial aggregation, Renewables
Article history: Received 17 September 2020; Accepted 2 January 2021; Available online 6 January 2021

DOI of original article: 10.1016/j.esr.2020.100563
E-mail address: christian.fleischer@uni-flensburg.de

https://doi.org/10.1016/j.mex.2021.101211
2215-0161/© 2021 The Author. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
Method details

Background

Cost optimisation results of energy system models with high penetration of solar and wind are impacted by choice of the spatial resolution of the model. Two documented effects are the increase in transmission expansion cost and the reduction in solar and wind capacity with increasing spatial resolution [18–20]. The max-p regions method presented in [17] can be effective in minimising these two effects by using several energy-related spatial attributes to define regions during spatial resolution reduction. The max-p regions method builds upon the approach applied by [1] and [2] to define regions. Similar to the clustering method used in [1], the max-p regions method uses multiple energy-related spatial attributes to define similarities between areas but the max-p regions method uses the max-p regions problem algorithm instead of k-means. The clustering algorithm used in [2] uses a combination of both the max-p-regions problem algorithm and k-means but [2] only one spatial attribute to define regions. Two clustering methods are used by [2] as solving the max-p regions problem algorithm as a mixed-integer programming problem with an increasing number of areas can become intractable. The presented max-p regions method uses a heuristic solution to solve the max-p-regions problem presented in [15]. The random regions method described in this article is used in [17] to verify the efficacy of the max-p regions method to define regions that are less impacted by spatial resolution reduction effects.

Spatial dataset preparation

A georeferenced dataset of the Nomenclature of Territorial Units for Statistics Level 2 NUTS 2 areas for countries in the European Union, plus Norway, Switzerland and the UK is used to apply the method. The spatial dataset is structured using the GeoDataFrame framework of the geopandas tool. The Identification (NUTS_ID), country reference code (CNTR_CODE), and the geometrical information (geometry) of the NUTS 2 areas, extracted from the Eurostat database [3], are added to the dataset. The coordinate reference system (CRS) of the geometry of the areas, is set as EPSG 3035.

Next, the population in the NUTS 2 areas is added to the dataset. The population values are extracted from the gridded population of the world version 4 (GPWv4) of 2015 [4], using the NUTS 2 area geometry and the zonal_stats function in the rasterstats python package [12]. The CRS of the GPWv4 GeoTIFF file is WGS84, and therefore the NUTS 2 areas geometry is converted to the same
The Geospatial Land Availability for Energy Systems framework (GLAES) [8] is the tool used to determine the eligible areas. The eligible land areas are calculated by applying the med scenario from [8] on the NUTS 2 area geometry. The eligible offshore areas within the exclusive economic zones are determined by eliminating non-eligible areas. These non-eligible areas are areas listed in the World Database on Protected Areas [7], areas with more than 60 m water depths, areas wherein 2017 ships were recorded to have spent in average more than one hour in a square km per month, areas within 12 nautical miles of the coast, areas within one nautical mile of gas and oil pipelines [6]. 30% of the remaining eligible area is proportionally distributed to NUTS 2 areas according to the size of their coastal borders to their exclusive economic zone. The installation capacity potential of offshore wind, onshore wind and solar are calculated using the capacity density values 5.36 MW/km², 4 MW/km² and 12 MW/km² respectively. The second factor used to determine the electricity generation potential, of wind and solar, is the average full load hours. The full load hours of technology is the annual sum of the normalised maximum power output profile assigned to the NUTS 2 area taken from the renewables.ninja platform [13]. The full load hours of the onshore wind and solar are multiplied with their respective potential installed capacities to calculate their electricity generation potential within a NUTS 2 area. The offshore wind electricity generation potential is calculated by multiplying the full load hours with the potential offshore capacity assigned to a NUTS 2 area. The technology with the highest electricity generation potential is assigned to the NUTS 2 area dataset.

### Grouping island NUTS 2 areas

Certain countries have NUTS 2 areas that are islands and thus do not share any vertexes with other NUTS 2 areas. The max-p-regions method has a contiguity constraint that ensures all NUTS 2 areas within a region shares at least one vertex with another NUTS 2 area in the same region. To enable that island NUTS 2 areas are capable of joining with other NUTS 2 areas to build regions they are grouped with the nearest continental NUTS 2 area. The geometries of these grouped NUTS 2 areas are joined to construct a single geometry. Table 1 has details of the islands and the continental NUTS 2 areas.

### Table 1

The island NUTS 2 areas and their respective continental NUTS 2 area to which they are attached.

| Groups | Island NUTS 2 areas | Continental NUTS 2 area |
|--------|---------------------|-------------------------|
| 1      | UKN0                | UKM3                    |
| 2      | ITG2, ITG1         | ITG3                    |
| 3      | FI20                | FI1B                    |
| 4      | DK01, DK02         | DK03                    |

The CRS before applying the zonal_stats function. The geometries of the NUTS 2 areas dataset is then transformed back to European Petroleum Survey Group (EPSG) 3035.
Pseudocode 1: MAX-P REGIONS METHOD

gdp: Geopandas dataframe of NUTS 2 areas without islands,
A: Set of areas in gdp,
W: Neighborhoods,
k: Spatially extensive attribute of the areas,
population: Population values in the areas,
maxVRE: Generation potential of solar or wind in the areas,
storage: Storage potential of pumped-storage hydropower in the areas,
d: Features to determine pairwise dissimilarities between areas,
Areas: Area value of the areas,
α: Area factor used to calculate threshold,
P: Partition indices defining the regions to which areas are assigned to.

\[
W = \text{weightsQueen}(gdp), \text{ weights object build using the pysal queen contiguity weights function.}
\]
\[
d = \{\text{population, storage, maxVRE}\}.
\]
\[
l = \text{Areas},
\]
\[
\eta = \text{Median of the area values in } A.
\]
\[
\text{threshold} = \eta \cdot \alpha.
\]
\[
\text{maxEnclave} = 10, \text{ maximum candidate regions in the enclave assignment process.}
\]
\[
\text{maxCon} = 999, \text{ maximum number of iteration for construction process.}
\]
\[
\text{maxSA} = 10, \text{ maximum number of iteration for simulated annealing process.}
\]
\[
\text{model} = \text{MaxPHHeuristic}(gdp, d, W, l, \text{threshold, maxEnclave, maxCon, maxSA}), \text{ build model heuristic max-p-regions problem.}
\]
\[
\text{Solve model and assign } P \text{ to the areas in } gdp.
\]

Fig. 1. Pseudocode detailing the parameters and functions used to conduct the max-p regions method.

Defining regions using the max-p-regions method

The Python Spatial Analysis Library (PySAL) contains a spatial optimisation library called spopt[14]. As described in the pseudocode 1 below, the MaxPHHeuristic function of the spopt library, in conjunction with the libpysal python library, is used to implement the max-p-regions method.

The MaxPHHeuristic function uses a heuristic approach of solving the max-p-regions problem which defines an objective function and constraints used to maximise the heterogeneity between the regions and maximise the number of regions created. As presented in [15], there are three options to conduct a local search when finding the best feasible solution to define regions. The MaxPHHeuristic function uses the simulate annealing approach. A contiguity constraint is defined in the MaxPHHeuristic function by assigning weight values to the NUTS 2 areas using the queen contiguity weights function provided by libpysal [16]. The max-p regions method is detailed in Fig. 1.

The result of the MaxPHHeuristic function returns an assigned partition value for each NUTS 2 area. The NUTS 2 areas that were assigned the same partition value are grouped into the same region. The size and number of regions created depend on the threshold value. By changing the area factor used to determine the threshold value, it is possible to vary the size and number of regions allows for the variation of the spatial scale of the energy system model. The area factor must be less than the fraction of the sum of the area values to the median of the area values of the NUTS 2 areas.

The energy-related max-p-regions method defines regions of NUTS 2 areas based on their energy-related spatial attributes. The aim of the method is to differentiate regions according to how they consume, produce and store electricity in a power system with high wind and solar energy penetration. Therefore three spatial attributes used to define heterogeneity of the NUTS 2 areas are population; wind and solar resource potential; and pumped-hydro storage capacity. These three
Example of creating three energy-related max-p regions and three random regions from the NUTS 2 areas of Germany. The energy-related attributes solar and wind potential, population and pumped-hydro storage are used to determine the heterogeneity between regions and create energy-related max-p regions. Random values from 0 to 1 are used to determine the heterogeneity between regions to create random regions.

The NUTS 2 areas spatial dataset can be filtered to focus on a particular set of NUTS 2 areas. The NUTS 2 areas of Germany were filtered to create energy-related regions and random regions depicted in Fig. 2.

An alternative method to define regions is to associate NUTS 2 areas at random to regions while maintaining the contiguity of the regions, hereafter referred to as the random-regions method. The random-regions method assigns a set of unique random values from 0 to 1 to the NUTS 2 areas. The random-regions method follows the pseudocode 1 in Fig. 1 but uses the random values instead of the energy-related spatial attributes to determine the heterogeneity between regions when creating the model with the MaxPHeuristic function. The random-regions method was used in [17] to investigate the effectiveness of the max-p regions method to minimise the effects of spatial resolution reduction on power system models. The NUTS 2 area spatial dataset can be filtered to focus on a particular set of NUTS 2 areas. The NUTS 2 areas of Germany were filtered to create energy-related regions and random regions depicted in Fig. 2.
The complete process to create the energy-related max-p regions and the random regions is illustrated in Fig. 3.

Conclusion

This article provides a detailed description of how to define regions using the max-p regions method. The method can help minimise the effects of spatial resolution reduction on an energy system model. The max-p regions method can be improved with more spatial data availability on the onshore and offshore area to more accurately determine the potential of wind and solar within the areas.

Declaration of Competing Interest

The author confirms that there are no conflicts of interest.

Acknowledgements

This work has been made possible through funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No: 765515.

References

[1] T. Anderski, Y. Surmann, S. Stemmer, N. Grisey, E. Momot, A.-C. Leger, B. Betraoui, P. van Roy, European cluster model of the Pan-European transmission grid - e-HIGHWAY 2050: Modular Development Plan of the Pan-European Transmission System 2050, (2015). https://docs.entsoe.eu/baltic-conf/bites/www.e-highway2050.eu/fileadmin/documents/Results/D2.2_European_cluster_model_of_the_Pan-European_transmission_grid_20072015.pdf (accessed March 13, 2019).

[2] K. Siala, M.Y. Mahfouz, Impact of the choice of regions on energy system models, Energy Strategy Rev 25 (2019) 75–85, doi: 10.1016/j.esr.2019.100362.

[3] Geographic Information System of the COMmission, GISC data distribution API - NUTS 2016, (n.d.). https://gisco-services.ec.europa.eu/distribution/v2/nuts/geojson/NUTS_RG_10M_2016_3035.LEVL_2.geojson (accessed July 7, 2020).

[4] Center for International Earth Science Information Network - CIRESN - Columbia UniversityGridded Population of the World, NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY, 2018 Version 4 (CPWv4): Population Count, Revision 11https://doi.org/https://doi.org/, doi: 10.7927/H4JW8BX5.

[5] JRC Hydro-power plants database, Contribute to energy-modelling-toolkit/hydro-power-database development by creating an account on GitHub, Energy Modell. Toolkit (2019), (accessed April 18, 2019). https://github.com/energy-modelling-toolkit/hydro-power-database.

[6] Pipelines | EMODnet Human Activities, (n.d.). http://www.emodnet-humanactivities.eu/search-results.php?dataname= Pipelines (accessed May 27, 2019).

[7] Protected Planet, Prot. Planet. (n.d.). https://protectedplanet.net/c/world-database-on-protected-areas (accessed May 27, 2019).

[8] D.S. Ryberg, M. Robinius, D. Stolten, Methodological Framework for Determining the Land Eligibility of Renewable Energy Sources, (n.d.) 35.

[9] S.J. Rey, L. Anselin, PySAL: a python library of spatial analytical methods, in: M.M. Fischer, A. Getis (Eds.), Handb. Appl. Spat. Anal., Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 175–193, doi:10.1007/978-3-642-03647-7_11.

[10] S. van der Walt, S.C. Colbert, G. Varoquaux, The NumPy array: a structure for efficient numerical computation, Comput. Sci. Eng. 13 (2011) 22–30, doi:10.1109/MCSE.2011.37.

[11] Kelsey Jordahl, Joris Van den Bossche, Jacob Wasserman, James McBride, Martin Fleischmann, Jeffrey Gerard, Jeff Tratner, Matthew Perry, Carson Farmer, Geir Arne Hjelle, Sean Gillies, Micah Cochran, Matt Bartos, Lucas Culbertson, Nick Eubank, Aleksy Bilogur, maxalbert, geopandas/geopandas: v0.7.0, Zenodo, 2020, doi: 10.5281/zenodo.3669853.

[12] rasterstats — rasterstats 0.10.3 documentation, (n.d.). https://pythonhosted.org/rasterstats/index.html (accessed July 7, 2020).

[13] S. Pfenninger, I. Staffell, Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data, Energy 114 (2016) 1251–1265, doi:10.1016/j.energy.2016.08.060.

[14] pysal/spopt, Python Spatial Analysis Library, 2020. https://github.com/pysal/spopt (accessed December 13, 2020).

[15] J.C. Duque, L. Anselin, S.J. Rey, The max-p-regions problem, J. Reg. Sci. 52 (2012) 397–419, doi:10.1111/j.1467-9787.2011. 00743.x.

[16] libpysal.weights. Queen — libpysal v4.3.0 Manual, (n.d.). https://pysal.org/libpysal/generated/libpysal.weights.Queen.html (accessed December 13, 2020).

[17] C.E. Fleischer, Minimising the effects of spatial scale reduction on power system models, Energy Strategy Rev 32 (2020) 100563, doi: 10.1016/j.esr.2020.100563.

[18] V. Krishnan, W. Cole, Evaluating the value of high spatial resolution in national capacity expansion models using ReEDS, in: 2016 IEEE Power Energy Soc. Gen. Meet. PESCM, Boston, MA, USA, IEEE, 2016, pp. 1–5, doi: 10.1109/PESCM.2016.7741996.
[19] J. Hörsch, T. Brown, The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios, in: 2017 14th Int. Conf. Eur. Energy Mark, EEM, 2017, pp. 1–7, doi:10.1109/EEM.2017.7982024.

[20] B.A. Frew, M.Z. Jacobson, Temporal and spatial tradeoffs in power system modeling with assumptions about storage: an application of the POWER model, Energy 117 (2016) 198–213, doi:10.1016/j.energy.2016.10.074.