EUROPA-UNIVERSITÄT FLENSBURG

DOCTORAL THESIS

The Role of Spatial Context in Energy System Models

Author:
Christian Etienne Fleischer

Supervisors:
Prof. Dr. Olav Hohmeyer
Prof. Dr. Bernd Möller

A thesis submitted in fulfillment of the requirements for the degree of Dr. rer. pol.

at the

Interdisciplinary Institute for Environmental, Social and Human Sciences
Department of Energy and Environmental Management

March 28, 2023

Dedicated to my family

Summary

The transition towards a sustainable energy system is an essential step towards mitigating GHG emissions and the effects of climate change. Energy system models are used to gain insights into different transition potential pathways. Various components of the energy systems with the inclusion of variable renewable energy technologies are temporally and spatially dependent on weather and climate conditions. With the improvement in the availability of highly spatially resolved data on energy-related attributes such as electricity demand and electricity generation potential, energy system models can be improved to better capture their variability. Given computational limits, the spatial resolution of energy system models is often reduced. The choice of spatial resolution reduction method or the data processing method used to generate the energy system models is usually not discussed or adequately documented. This dissertation presents a documented data processing approach that integrates different spatial resolution reduction methods.

There are two documented effects of spatial resolution reduction of energy system models with high wind and solar PV penetration. The two effects are the loss of good solar and wind sites and the reduction in transmission capacity expansion. This dissertation presents and investigates the use of a targeted novel resolution reduction method named the max-p regions method to reduce the spatial resolution of energy system models. The max-p regions method reduces the spatial resolution using the max-p regions problem algorithm and three energy-related spatial attributes. The spatial attributes are electricity consumption, solar and wind generation potential and energy storage capabilities. A comparative analysis was conducted to evaluate the effectiveness of the max-p regions in minimising the impact of two documented spatial resolution reduction effects. The results of the evaluation show that the power system models of two out of the four European countries investigated was less impacted by the effects of spatial resolution when the max-p regions method was used to define the regions of the models in comparison to the use of a non-targeted method of defining regions. The use of the max-p regions method to define regions in sub-continental European power system models was impacted less by the effects of spatial resolution reduction compared to the use of national jurisdictions.

Publication List

This thesis is based on the following publications:

A Christian Etienne Fleischer. Minimising the effects of spatial scale reduction on power system models. en. In: *Energy Strategy Reviews* 32 (Nov. 2020), p. 100563.

B Christian Etienne Fleischer. Using the max-p regions problem algorithm to define regions for energy system modelling. en. In: *MethodsX* 8 (Jan. 2021), p. 101211.

C Christian Etienne Fleischer. A data processing approach with built-in spatial resolution reduction methods to construct energy system models. en. In: *Open Research Europe* 1 (Feb. 2022), p. 36.

Acknowledgements

I want to thank all those who have provided me with support and guidance since I started my PhD. I want to first express my appreciation to Prof. Olav Hohmeyer, who has encouraged me to undertake this PhD. His extensive understanding of the energy field has been instrumental to me finding and better comprehending the context of my research focus. I am also grateful to Prof. Müller for sharing his expertise in the field of heat modelling. Furthermore, I would like to thank my colleagues at the Centre for Sustainable Energy Systems that have assisted me with my work and with whom I have shared insightful discussions over coffee and lunch breaks. I would like to thank Maruf and André for their camaraderie both in and outside of the workplace.

I want to thank the ENSYSTRA project group for the training and the exposure to different disciplines involved in the energy system transition. Interacting with the participants has been an excellent learning and fun experience. Having constant access to a network of talented early-stage researchers has been very helpful and encouraging.

Last but not least, I would like to thank my family for without their support, this undertaking would not have been possible. My family members have been my role models, as many of them have excelled in their endeavours, paving the way for others to follow. A special mention goes to my wife, Annabelle, whose love, wisdom and encouragement have been a constant in my life.

Contents

List of Abbreviations

ASHP	Air-source heat pump
CAES	Compressed air energy storage
CO_2	Carbon dioxide
COP	Coefficient of performance
EU	European Union
FERC	Federal energy regulatory commission
GDP	Gross domestic product
GHG	Greenhouse gas
GIS	Geographic information system
GW	Gigawatt
HDV	Heavy duty vehicle
LDV	Light duty vehicle
LE	Land eligibility
LULUCF	Land use, land-use change and forestry
MILP	Mixed integer linear program
NUTS	Nomenclature of territorial units for statistics
OSM	Open street map
PV	Photovoltaic
RPS	Renewable portfolio standard
TSO	Transmission system operator
VGI	Volunteered geographic information
VRE	Variable renewable energy

Part I: The Role of Spatial Context in Energy System Models

Introduction

The EU is one of the economies that have contributed the most towards climate change through their anthropogenic GHG emissions. It is estimated that based on historic cumulative GHG emissions before 2012, the EU28 countries are responsible for 17.3%, the second-highest contributor after the US, of global temperature increase in 2100 [1]. The governments, including those on the European continent, have adopted and ratified the Paris Agreement, which states that all parties should strive towards keeping the increase in global average temperature "well below" 2°C while trying to maintain it at 1.5°C above pre-industrial levels. With an aim to take the lead in climate action, the EU commission has a 2030 Climate Target Plan that aims to reduce 55% GHG emissions by 2030, while the EU parliament has announced a 60% reduction target by 2030. These proposals support the long-term EU strategy of reaching climate neutrality by 2050. The EU plans to legally bind the climate targets and long-term strategy for all member states in a new European Climate Law. According to Climate Analytics, only a minimum of 65% reduction in GHG emissions would ensure that the EU27 countries are compatible with keeping the increase in global average temperature to 1.5°C above pre-industrial levels [1]. In 2018, the EU recorded a GHG emission reduction of 25.2% since 1990 (without LULUCF).

As summarised in the IPCC special report by Rogelj et al. [2], pathways consistent with the Paris Agreement require rapid phase-out of CO_2 emissions and deep emissions reductions in the energy, industry, transport, buildings, agriculture, forestry and other land-use sectors. Among the characteristics observed by Rogelj et al. [2] in 1.5°C consistent pathways is the rapid decrease in the carbon intensity of the electricity generation and simultaneous increase in electrification of energy end-use. The carbon intensity of electricity generation can be decreased by increasing renewable energy technologies in the energy generation mix. The power, heat and transport sectors are the most significant GHG emitting sectors in the EU and need to be decarbonised to achieve the EU carbon neutrality target. Road transport containing mainly LDV and HDV represented 26% of overall CO_2 emissions within the EU-28 and Iceland in 2017 [3]. It is expected that there will be an increase in the use of electric mobility and alternative energy carriers to decarbonise the road transport sector. Examples of alternative energy carriers are hydrogen and synthetic hydrocarbons generated from power to gas technologies. The public electricity and heat generation represented 27% of overall CO_2 emissions within the EU-28 and Iceland in 2017 [4]. There has been significant progress in the power and heat sector in the EU with the adoption of deployment of renewables and the application of energy efficiency measures. Renewable technologies such as wind and solar PV will continue to play an essential role in achieving the decarbonisation of the power, heat and transport sector. Their share in the electricity generation mix will have to continuously increase to achieve the climate targets.

The transition away from the current energy system regime dominated by fossil fuels brings multiple opportunities and challenges that lead to different pathways. One challenge of the continuous increase in wind power and solar PV in the EU energy mix is the increased penetration levels of variable electricity generation in the power system. These variations are temporally and spatially dependent on local weather conditions and need to be managed to prevent adverse effects on the reliability of the power systems. There are different methods of managing the variations, which opens opportunities to management strategies such as the use of storages technologies, sector coupling and grid expansion. Storage technologies options can range from pumped storage to batteries installed in electric vehicles. Whereas sector coupling opportunities such as using power to gas technologies to replace fossil fuels as heating and transport fuels can also help manage supply and demand variability.

The transition towards a sustainable energy system can have different narratives with different configurations of technology deployment over time. Energy system models are used to analyse these different narratives and their respective potential benefits and drawbacks.

1.1 Research Interest

This dissertation aims to improve the current understanding of how the spatial context of energy system components impacts energy system models with the high penetration of renewables. The spatial context considers the choice of spatial aggregation and the choice of geographical scale. This dissertation also aims at developing new methods and tools that allow for better representation of spatial data in energy system models.

Together the publications presented and discussed in this dissertation address the overall aim of this dissertation. The specific aims of the individual publications are:

- To investigate how the effects of spatial scale resolution reduction can be minimised by proposing the max-p regions method, which aggregates areas according to the similarity in energy-related spatial attributes (**Paper A**).

- To define the methodological details of a spatial resolution method that minimises the impact of spatial scale reduction when defining regions for the energy system model (**Paper B**).

- To present a data-processing approach that integrates spatial data and spatial resolution reduction methods in energy system modelling. Investigating the impact of the choice of the method applied to define regions and the choice in geographical scale considered on the results of an energy system model (**Paper C**).

1.2 Research context

This dissertation research focuses on the European energy system and how the effects of spatial data can impact the description and results of energy system models used to analyse the transitions towards a sustainable energy system. Building energy system models for continental or even national scale often can lead to models having complexity levels that exceed the existing computational limits. Therefore these large-scale model complexities are reduced spatially, temporarily or in some cases even technologically. This dissertation focuses on the spatial complexity reduction of the energy system models. More specifically, the spatial resolution of energy system components is reduced at levels that allow for computational tractability of energy system models. This dissertation, therefore, analyses the benefits and limitations of using different spatial resolution reduction methods.

Several case studies are used to compare the spatial resolution reduction methods and their effects on energy system models. These case studies include single European countries and groups of European countries. The single European countries include Germany, Spain, Italy and Great Britain. Although the case studies are limited to the European context, the body of literature used in this dissertation is not.

The energy sectors that are considered in this dissertation are the power and heat sector. The choice of sectors is determined based on the available information needed to represent them in energy system models adequately. The heat sector is also considered in this dissertation as the interconnection and interdependency of the power sector, and the heat sector is continuously increasing. Considering the heat sector and the power sector together in energy system models also allows for a better understanding of interactions and synergies between the two sectors.

The research in this dissertation was conducted using open science practices. Therefore the energy system models used were build using the open energy modelling process described by Pfenninger et al. [5]. The data and the code used to formulate the model have been made accessible and legally usable. The publications used in this dissertation are also openly available under licences that permit unrestricted use, distribution, and reproduction of the article. By using open science practices, the work conducted aims at promoting transparency and reproducibility of model research results. This work aims at limiting duplication of work by using existing openly available datasets to build the models used in the research conducted.

1.3 Structure of the thesis

The first part of this dissertation presents an overview of the research conducted to address the overall research aim. The structure of the first part of this dissertation begins with the introduction section, followed by a background section that establishes state of the art in this field of research. The methodology section describes the methods and scenarios used to build and analyse energy system models. The results are presented and discussed, and then the final section concludes the findings of the dissertation.

The second part of this dissertation is a collection of the following publications:

Paper A is a peer-reviewed journal paper in *Energy Strategy Reviews*. This publication analyses the efficacy of a proposed max-p regions method in minimising two effects of spatial resolution reduction. The two documented effects are the loss of good solar and wind sites and the reduction in transmission bottlenecks.

Paper B is a peer-reviewed method paper in *Method X*. This publication gives a detailed description of how energy system models can use a spatial resolution method, named the max-p regions method, to join areas into regions with similar energy-related spatial attributes.

Paper C is a research article in the *Open Research Europe* publishing platform. This publication focuses on providing a detailed description of a data-processing method using datasets that includes GIS data. This data-processing method also integrates different spatial resolution reduction methods to define regions when formulating energy system models. A case study is presented in the publication to demonstrate the impacts of spatial scale and spatial resolution reduction methods on energy system optimisation results.

Background

In this section the use of spatial data to inform energy system models is introduced. First, the spatial sensitivities of energy system components in the power and heat sector are discussed. Next, the state of the art approaches of using open source software to integrate geospatial information in data processing approaches to build energy system models is presented. Finally, the documented effects of spatial resolution reduction are reviewed.

2.1 The spatial sensitivities in energy system models

Traditionally geospatial data was less relevant in the formulation of power system models as traditional energy systems were structured around centralised dispatchable generators. Today power systems are transitioning towards a more decentralised topology dominated by VREs such as wind and solar. The capacity factors of VREs are spatio-temporal specific. Thus, to adequately model power sectors with high shares of VREs requires power system models to have a sufficient spatial and temporal resolution [6]. The following sections look at the spatial and temporal modelling of power system components.

2.1.1 Variable renewable energy

Site-specific parameters such as direct and diffuse irradiation, in the case of solar PV, determine the performance and cost of VREs. Various tools make use of large datasets, which can include satellite-based climate data or reanalysis data to calculate the spatiotemporally resolved data on the availability of VREs. The renewables.ninja platform, for example, provides historical values for power outputs of solar PV, onshore wind and offshore wind. The platform uses meteorological reanalysis, satellite-measured data and measured data from PV sites to generate the power output values. The highest spatial resolution of the power output values for solar PV and onshore wind on the platform is NUTS 2 level for the countries on the European continent. Other tools such as PVlib, PyGreta and Alite offer modellers more flexibility to calculate power values of VREs for specific geographical areas or even customised technical parameters. When using these tools, the spatial resolution of the power output of the VREs is only limited by the resolution of the satellite weather data or the reanalysis of weather data applied. For example, Alite allows for the creation of power output of solar PV systems using ECMWFs' ERA5 reanalysis dataset, which has a spatial resolution of ca. 30 km x 30 km [7].

A key differentiating characteristic of solar PV onshore wind and offshore wind is that they require relatively larger areas per installed capacity than conventional power plants. Spatial analysis is undertaken to determine the eligible areas for VREs installations and the total capacity that can be installed given the specifications of the technology. LE analyses are undertaken and used as input parameters in energy system models to define the maximal capacity of solar PV and onshore wind that can be installed in a specific region. Proximity to settlement, infrastructure, protected areas, and agriculture are commonly used criteria used in an LE analysis. Ryberg, Robinius, and Stolten [8] developed the GLAES tool and used European countries to demonstrate how the impact of LE constraints are highly spatially sensitive. Ryberg et al. [9] show that the LE for onshore wind is sensitive to socio-technical criteria such as minimal wind speeds, the maximal terrain slope, the maximal distance from power lines, and the minimal distance from settlements. Therefore it is crucial that the method applied to determine LE in energy system models must be clearly defined.

2.1.2 Demand profiles

Power demand profiles have spatial-temporal characteristics that are determined based on parameters such as population density, building heating specifications, temperature and human behaviour. The focus of energy demand modelling has historically been on the national scale or the building level [10]. Spatial load forecasting approaches are used for network planning which determines the placement of network components such as substations and feeders [11]. Limei Zhou et al. [12]

introduce a method that uses GIS to conduct spatial load forecasting to provide the growth of future load demand in space.

Access to demand data at high spatial resolution is often lacking, and therefore top-down approaches are used to disaggregate national scale data using socioeconomic data. Socioeconomic attributes often used to spatially disaggregate power profiles are GDP and population. Other top-down approaches also include using regression models and satellite data on nighttime lights as applied by Pan and Li [13] to generate spatial maps of power consumption for China. Mattsson et al. [14] have integrated machine learning in combination with other parameters to generate synthetic power profiles at spatial scales lower than the national level.

2.1.3 Energy storage

One method of balancing electricity demand and supply is the use of energy storage options. An established large-scale, cost-effective option of rechargeable energy storage is pumped-hydro storage. Pumped-hydro storage uses the height difference of reservoirs to store, consume and generate electricity. Data availability on the existing pumped storage installation across Europe is still not entirely freely accessible. Though incomplete, the JRC published a georeferenced open database containing 4131 hydropower installation across Europe, including pumped-storage hydropower plants [15]. Gimeno-Gutiérrez and Lacal-Arántegui [16] uses a GIS-based method and reservoir datasets to assess the potential for pumped hydropower in 14 European countries and found that the existing capacity could increase by 3.5 times to reach 54 TWh. The two other types of hydropower technologies are reservoir-based hydropower and run-off river hydropower, which depend on the inflow of water determined by rainfall and the melting of snow. The suitability of the topology requirements and the inflow of water makes these hydropower technologies spatially specific. CAES is another storage option where the geological properties of an area play an important role as suitable existing storage caverns makes this technology significantly cheaper [17].

2.1.4 Transmission

Transmission networks, which can transfer large units of electricity across large distances, is another method of balancing electricity demand with electricity supply. The expansion of transmission network can be a cost-effective way of integrating more renewables in the energy system. Schlachtberger et al. [18] show that using expanding transmission networks on a continental scale to balance the fluctuations of solar and wind in space is more cost-effective than using storage. The use of transmission networks maximises the of uses synoptic-scale weather differences to balance the energy supply and demand. Schaber et al. [19] calculate how backup capacity requirements and overproduction are reduced with transmission capacity expansion in highly renewable power system. Schaber et al. [19] also show that the expansion of the transmission network allows for the spreads of the burden of balancing the VRE supply across connected regions using flexible generators and storage options such as pumped-hydro storage. Spatial information on the electrical grid has been unavailable to the public in the past, but access has improved in the last decade . OSM is a VGI project that has mapped the electrical grid using crowd-sourced data [20]. Various open-source tools use different approaches to simulate grid models using OSM data that can then be used in energy system models.

2.1.5 Heat sector

As stated by Novosel et al. [21], heat has a key differentiation to electricity in that it cannot be transported over relatively large distances and therefore needs to be consumed within close confinement of where it is produced. District heating networks offer a heat supply strategy that can transport heat from sources such as excess heat from industry and power generation to meet the local heat demand. The assessment of the technical and economic potential of district heating in an area requires a spatial understanding of the heat demand densities. The use of heat atlases describing the demand densities at 100 m resolution is used by Möller et al. [22] to calculate the investment costs of heat distribution networks.

2.2 Integration of GIS in energy system models

As discussed by Martínez-Gordón et al. [23], there is no clear consensus on how to link GIS and energy tools. In the analysis of 34 open-source energy system models, Martínez-Gordón et al. [23] identified four models with an internal GIS-related function and five other models linked to a GIS tool. The majority of the energy system models use the internal GIS function to determine resource potential such as biomass and VRE potentials based on weather data. PyPSA (Hörsch et al. [24]) is one of the energy system modelling frameworks with an internal GIS function used to reduce the spatial resolution of the energy system model.

Modelling frameworks such as PyPSA, calliope [25] and oemof [26] allow for the use of GIS coordinates to identify the location of energy system components. The GIS coordinates of the centroid of the regions often represent the location of the regions in the energy system models. Except for the proposed data-processing method detailed in Paper C, the author is aware of two data processing approaches, PYSA-Eur [24] and euro-calliope [27], that have integrated spatial clustering functions to alter the spatial granularity of energy system models.

2.3 Effects of spatial resolution reduction methods

There are two well-documented effects of spatial resolution reduction. The first effect of spatial resolution reduction is the decrease in transmission bottlenecks with decreasing the spatial resolution of energy system models. Hörsch and Brown [28] investigated the impacts of spatial scale on a highly renewable European power system. The spatial scale was varied by varying the number of clusters of the power system from 37 clusters to 362. An increase in transmission capacity expansion cost was observed when the spatial resolution of the model was increased due to higher line capacity and line constraints. These line constraints, which manifest as transmission bottlenecks, prevent electricity generated from good wind sites to travel to load centres. In Germany, Hörsch and Brown [28] observed a reduction of wind capacity from 40 GW to 12 GW and solar PV capacity increasing from 46 GW to 100 GW when the spatial resolution was increased from 37 clusters to 362 clusters as a result of transmission bottlenecks. Shawhan et al. [29] analysed the effects of simplifying the transmission model of eastern US and Canada on carbon dioxide emission abatement when imposing a carbon dioxide emission allowance price within certain areas of the grid. They found that a more simplified grid had fewer transmission constraints, allowing for more electricity importation from areas where the emission price was not in effect. Frew and Jacobson [30] investigated the effects of spatial scale of wind and solar site development on the power system models with high shares of renewables in the US. They observed that not considering intra-region transmission lines favoured the adoption of large-scale solar PV but did not favour residential solar PV.

The second effect is the increase in solar and wind capacity expansion cost with decreasing spatial resolution reduction of energy system models. This effect is the cause of the loss of good solar and wind sites with decreasing spatial resolution. The loss of good sites is the result of aggregating low-yield sites together with high-yield sites. Therefore there is a loss in the opportunity to maximise the use of high-yield sites, which increases the investment and cost of solar and wind installations. Frysztacki et al. [31] show that increasing the spatial resolution of a European power system model with a high share of renewables from 37 nodes to 1024 nodes reduces the system cost up to 10%. Frew and Jacobson [30] found similar system cost reductions of up to 42% in the system cost with an increased spatial resolution by modelling individual wind and solar sites instead of uniform buildouts. The AllCA region, which is a collection of FERC regions in California, quantified a 10% reduction in system cost and a 20% reduction in overgeneration at 100% RPS target [30]. Krishnan and Cole [32] found that in addition to a decrease in system cost, the deployment of solar PV decreased, and wind deployment increased with decreasing the spatial resolution of the US power system. Hörsch and Brown [28] observed a reduction in overall system cost with higher spatial resolution attributed partly towards better exploitation of good wind and solar sites. An example of the loss of good wind site is observed in Germany, where the best wind site in the power system model with 362 clusters has a weighted average capacity factor of about 40% in comparison to the single node representation of Germany, which had the weighted capacity average of only 26% [28].

The effects of spatial scale reduction on wind generation were even observed at the geographical scale of Austria by Simoes et al. [33]. Simoes et al. [33] discuss that complementary to the loss of good sites other mechanisms affect energy system models in relation to renewable energy generation when there is a reduction in spatial scale. These mechanisms are the region-specific costs, suitability of regional wind and solar profile to fit with the demand profile, and distance between electricity generation and demand locations impact the results of the energy system model at various spatial scales.

The two main effects of spatial resolution reduction have knock-on impacts on the results of energy system models as certain technologies complement each other. For example, as documented above, Hörsch and Brown [28] observed a reduction in offshore wind development when the spatial resolution is increased due to more intra-country transmission bottlenecks and improvement of capacity factors of onshore wind close to load centres. These two effects had the knock-on effect of more adoption of solar PV, which increased battery capacities and decreased hydrogen storage as batteries can smooth out the diurnal effects of solar PV at a lower cost in comparison to hydrogen storage.

2.4 Impact of spatial extent

As investigated by Tröndle et al. [34], the geographical scale of a fully renewable energy system impacts the deployment of supply, transmission and balancing technologies. Tröndle et al. used a European powers system model at 497 subnational regions to show that if the supply and balancing of power are limited at the regional or national level, then the overall system cost increases by 1.4 and 1.69 times the most cost-effective system, respectively. The most cost-effective system allows for balancing and supply of electricity at a continental scale. The main increase cost drivers are given as the limited access of certain sub-national regions to balancing options such as hydropower and the inability to share the best solar and wind resources. Tröndle et al. [34] also presented that certain subnational regions generated more than 50 times their power demand at continental scale supply and demand, meaning that much of the supply of power is unequally spatially distributed and that certain regions use more of their land area for wind and solar PV deployment than others.

Frew and Jacobson [30] made a similar observation to Tröndle et al. [34] while investigating the effects of spatial extent on a power system model of the US. Frew and Jacobson found that the interconnected US power system was more cost-effective at reaching an 80% RPS target than when the individual FERC regions were left to achieve the same target independently. The interconnected US power system model prioritises areas in certain regions with good wind sites in the Great Plains, Rocky Mountains, and Pacific Northwest. Frew and Jacobson [30] also observed a disparity in the contribution of the regions in achieving the 80% RPS target when the regions were interconnected. While certain regions exceeded the target to become fully renewable, others only reached as low as 56% RPS.

The importance of spatial scale in an interconnected European power system in relation to minimising the variability of wind power output is investigated by Grams et al. [35]. Their study observes that multi-day volatilities present themselves as electricity surpluses and deficit from wind installations lasting days to weeks in certain European regions. These multi-day volatilities could be addressed by strategic deployment of wind installations in the peripheral regions as these regions have contrasting wind regime behaviours to wind regimes in regions like the North Sea region. Addressing multi-day volatilities in wind power output helps maintain mean generation and reduces the need for flexibility options such as storages and flexible demand. Santos-Alamillos et al. [36] were able to model a substantial reduction in wind power output fluctuations in the southern Iberian Peninsula by taking advantage of the spatial variability of the wind energy resource in that region.

Analysing an energy system at different geographical scale raises the question of energy autarky. Tröndle, Pfenninger, and Lilliestam [37] investigate autarky at four different geographical scales (continental, national, regional and municipal level) in a European renewable-based power system. Tröndle, Pfenninger, and Lilliestam [37] found that the electricity demand in certain regions and municipalities exceeds the technical-socio potential of solar PV and wind. The analysis of energy systems at various spatial extent provides insights into autarky and energy security policy implications.

2.5 Spatial resolution reduction methods

The spatial resolution reduction methods of large scale energy system models can be classified into three categories. The first category uses spatial attributes such as political boundaries or market boundaries [23]. The use of political boundaries such as national jurisdictions to define regions in energy system models is commonly used as often the input data for the model is available only at that spatial resolution. With the increase in energy system input data at higher spatial resolution, new methods of spatial resolution reduction methods have emerged, which are classified in the second and third category. The second category, hereafter referred to as power system attribute-based clustering methods, are methods that use clustering algorithms and power system attributes to reduce the number of nodes in a highly spatially defined power system model. The third category uses spatial attributes in combination with clustering algorithms to build homogeneous regions. The third category is hereafter referred to as spatial attribute-based clustering methods. The second and third category are elaborated on in the following sections.

2.5.1 Power system attribute-based clustering methods

Hörsch and Brown [28] present a method tailored towards the co-optimisation of generation and transmission expansion investment which uses the k-means clustering algorithm to reduce the number of nodes in their European power system model. The k-means algorithm uses power system georeferenced data such as average power demand and conventional capacity values at substations. A benefit of this method is that it aims at retaining the topology of the current electrical grid. The same method is used by Müller et al. [38] to reduce the German power system model from 11,305 buses to 300 buses. The same method was adopted by Frysztacki et al. [31] to conduct their analysis of network resolution on power system models.

It must be noted that before the publication of the three studies mentioned above, other studies used clustering algorithms, including the k-means method, to reduce the complexity of a power system model dating as far as 1994 [39].

2.5.2 Spatial attribute-based clustering methods

To build a simplified European grid model, Anderski et al. [40] define 106 geographical clusters, which included the use of a clustering method and a combination of spatial attributes. The grid model was developed to be used in system simulations undertaken in the e-Highway project. Anderski et al. [40] used multiple steps to define geographical clusters. The first step used a combination of the k-means algorithm and the Tabu-search algorithm from the ClusterPy library [41] to find homogenous clusters in each European country. The clustering process, therefore, was conducted in a manner to prevent clusters from containing areas of different countries. The objective function of the clustering algorithm uses weighted spatial attributes to define the heterogeneity of NUTS 3 areas.

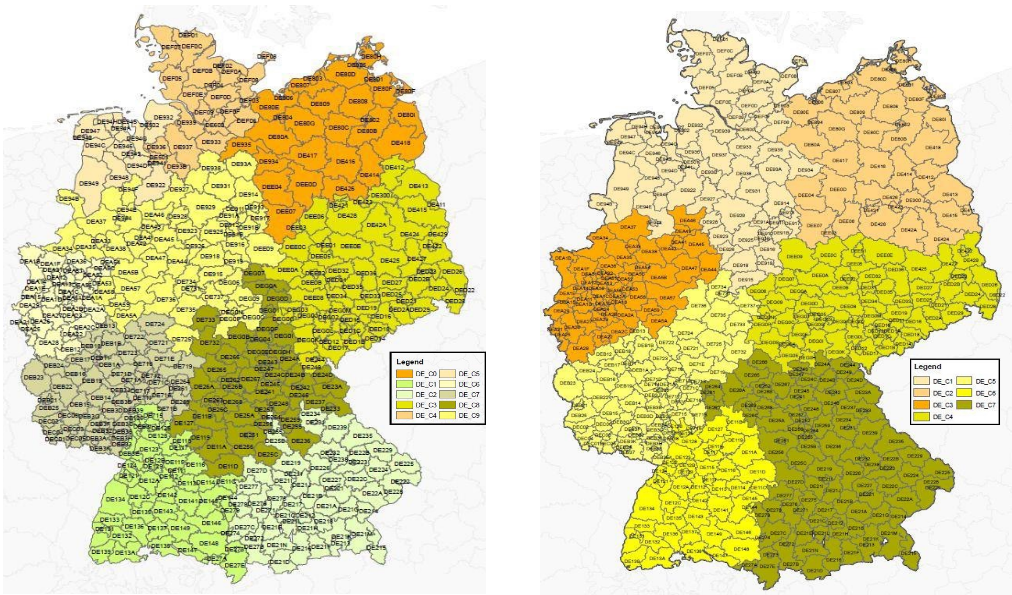


FIGURE 2.1: Clusters of NUTS 3 regions in Germany defined using clustering method (left) and after consultation (right). Adapted from Anderski et al..

The spatial attributes selected by Anderski et al. [40] are attributes that reflect the energy demand and supply potential in a NUTS 3 area. These include projected population, historical wind speed and solar irradiation average. The weights used in the objective function ranged from one to three and were assigned to the spatial attributes according to their significance to determine the clusters. For example, agriculture areas and natural grassland were assigned the lowest weight value while hydro installed capacity is assigned the highest weight value as power exchanges between geographical clusters and those without hydro capacity are expected to be significant. The clustering method applied does not ensure that the clusters generated have contiguous shapes, therefore, the latitude and longitude values of the NUTS 3 areas are also used as an attribute considered in the objective function. The second step implemented to define the geographical clusters used by Anderski et al. was a consolation process with the TSOs to adjust the clusters to account for consideration not considered by the clustering process, such as grid constraints.

Siala and Mahfouz [42] present the second application of clustering algorithms in combination with spatial attributes. The clustering method aims to aggregate high spatial resolution data of load density distribution, solar and wind potential with more than 108 data points to 28 contiguous regions. The raster data is first aggregated for each spatial attribute using an enhanced version of the k-means algorithm called k-means++. The clusters generated in the first step are then regionalised to lower spatial resolutions using the max-p regions method, which ensures that the clusters have contiguous shapes. The method uses the k-means ability to cluster large datasets and the ability of the max-p regions problem algorithm to build contiguous clusters. The max-p regions problem algorithm was introduced by She, Duque, and Ye [43].

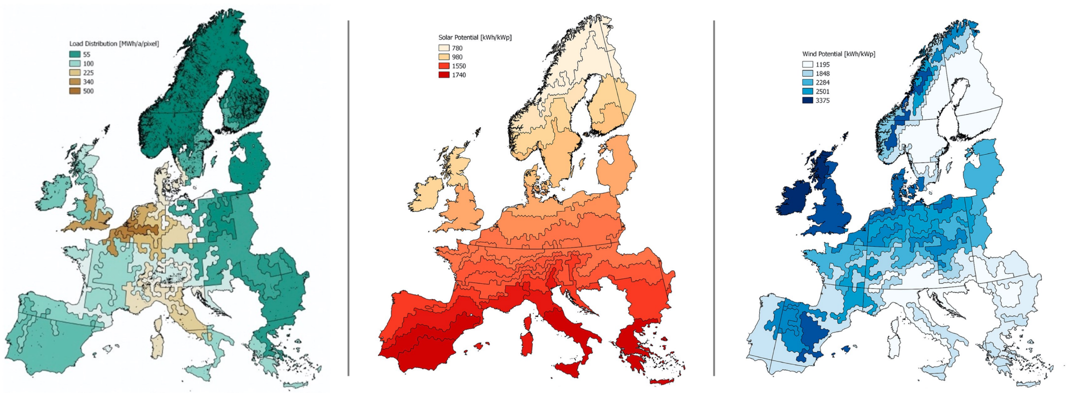


FIGURE 2.2: Clusters in Europe defined using clustering method and based on load density (left), solar potential (centre) and wind potential (right). Adapted from Siala and Mahfouz.

Getman et al. [44] evaluate the use of k-means and the max-p regions clustering algorithm to reduce the spatial resolution of solar PV capacity factor profiles in Colorado, US. The two methods are evaluated on how similar the data within a cluster is to the data representing the whole cluster. The two clustering methods are tested first on the measure of variance within the clusters by calculating their sum of squares and then using a cross-validation within each cluster. Getman et al. [44] conclude that the max-p algorithm generates the best results in terms of generating clusters that have a reasonable approximation of the data within the clusters.

Methodology: Models, scenarios, case studies and spatial resolution reduction methods

This section details the methods used to create models and scenarios to analyse the impacts of spatial context on energy system models in **Paper A** and **Paper C**. The novel data processing approach used to develop sector coupled energy system models introduced in **Paper C** is amongst the methods described. The spatial resolution reduction methods integrated into the data processing approach, including the max-p regions method presented in **Paper B**, are also presented in this section.

3.1 Models and scenarios

Optimisation models were used to investigate the effects of spatial context on energy system models with high penetration of renewable energy generation. The model structure, the data processing approach, the energy sectors considered, and the modelling framework used to formulate the models differ between **Paper A** and **Paper C**. A brief description and key assumptions of the models in the two papers are given in the following sections.

3.1.1 Power system optimisation models

The power system optimisation models used in **Paper A** are linear optimisation models with high penetration of renewables. The PyPSA modelling framework, is used to formulate power system optimisation models. PyPSA was selected as it is a modelling software tailored for power system modelling. The objective function of the optimisation models minimises the investment and dispatch cost of the power system, as described in Equation 3.1. The power system components considered in the objective function include generators, storages and transmission lines. The optimised models need to ensure that, in all instances, the power demand is met. The optimisation models are optimised for a period of one non-leap year at a three-hourly time resolution, which is a total of 2920 instances. A three hourly time resolution is used instead of hourly time resolution to improve the trackability of the models.

$$\min_{G_{n,s},F_l,g_{n,s,t}} [\sum_{n,s} c_{n,s} \cdot G_{n,s} + \sum_{n,s,t} c_l \cdot F_l + \sum_{n,s,t} o_{n,s} \cdot g_{n,s,t}] \tag{3.1}$$

where:

- n region;
- s generation or storage technology;
- l transmission line;
- t time interval;
- $c_{n,s}$ investment cost per installed capacity of generation and storage technologies;
- $G_{n,s}$ generation and storage technologies installed capacity;
- c_l transmission lines investment cost per line rated capacity;
- F_l transmission lines rated capacity;
- $o_{n,s}$ generation and storage technologies variable costs per unit electricity dispatched;
- $g_{n,s,t}$ dispatched electricity from generation and storage technologies.

An optimisation model was formulated for each geographic area and the spatial resolution reduction methods listed in Table 1. The following sections discuss the main assumptions used during the formulation of the power system optimisation models and the case studies modelled.

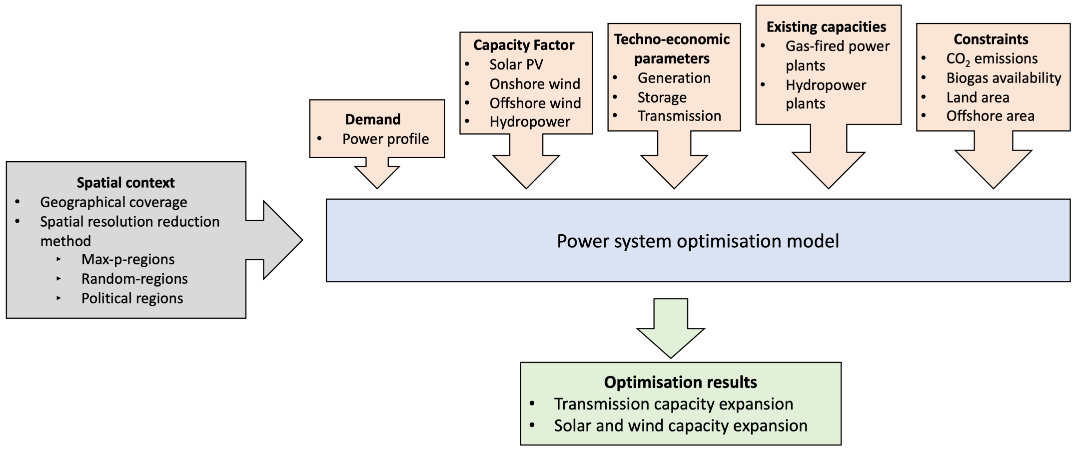


FIGURE 3.1: Illustrated description of the power system optimisation models formulation components.

A set of key assumptions were used to define the overarching scenario applied to the power system optimisation models, as indicated in Figure 3.1. A CO_2 emissions budget constraint is applied to the models to ensure the high penetration of renewables in the optimisation results. The CO_2 emissions budget limits the emissions of power generated by the use of natural gas in the model. A collective responsibly approach by European countries is assumed to calculate the respective natural gas budget for the different case studies modelled. The CO_2 emissions budget constraint limits the collective CO_2 equivalent emissions of the power sector of EU27 plus Great Britain, Norway and Switzerland to 5% of their 1990 CO_2 emissions. The operation of gas-fired power plants in the model is constrained by the natural gas emission budget and the availability of biogas. Like natural gas, biogas is a commodity collectively shared between the EU 28 countries plus Norway and Switzerland. The annual budget for biogas availability is an estimated amount of biogas utilised for electricity production in 2015. A conservative approach is taken to assume that biogas production for electricity consumption is not increased. Firstly, this assumption considers that biogas production has to compete with other uses of land area, which is limited. Secondly, the conservative approach accounts for the fact that other sectors, such as the heat and transport sector, will compete with the power sector for the limited amount of biogas to meet the EU climate targets.

Another constraint is the maximum allowable installed capacity of renewables in the models. Solar PV and onshore wind compete for the limited eligible land area, whereas onshore wind is limited by the eligible offshore area available. Both area availability constraints are set for each region in the model.

The optimisation models are not greenfield models as they are forced to install existing capacity of run-of-river hydropower, reservoir-base hydropower, pumped-storage hydropower and gas-fired power plants found within the regions of a model. The only rechargeable storage technologies in the model are pumped-storage hydropower, lithium-ion battery and hydrogen. It is important to note that only battery and hydrogen can increase their storage and dispatch capacity. Not allowing pumped-storage hydropower to expand is a limitation of the model and does not reflect the potential growth in capacity for pumped-storage hydropower in Europe. However, the purpose of this limitation is to prevent certain models from becoming too complex and unsolvable.

To ensure homogeneity of the power profile with the capacity factor of the variable renewable energy sources, the same reference year, 2010, is used for both input data sources. Considering that the power system models are considered potential simplified near future power systems, projected values for the year 2030 is used for investment and dispatch cost of technologies. The investment cost is annualised to account for their difference in lifespan.

3.1.2 Power and heat optimisation models

The power and heat optimisation models used in **Paper C** expand on the power optimisation model to include the residential and tertiary heat sector. The heat demand considers two end-uses of heat in the two sectors, which are hot water and space heating demand. A novel data processing approach was developed to build the power and heat optimisation models used in **Paper C** . The following sections provide more detail to the novel data processing approach, the heat demand modelling, and the case studies.

Data processing approach

The data processing approach was built based on three guiding principles. The first principle is to limit duplication of work by maximising existing openly available web-hosted pre-processed input data. The second principle is to incorporate multiple spatial resolution reduction methods within the data processing approach. The final principle is to encourage homogeneity between the availability of variable renewable energy generators with the power and heat demand.

The data processing approach is structured in two steps. Step one is the construction of a dataset called the Areas dataset. This dataset contains the data variables needed to build a power and heat optimisation model at NUTS 2 spatial resolution. The Areas dataset contains data for the 27 EU countries plus Great Britain, Norway and Switzerland. The data variables are primarily obtained from existing web-hosted pre-processed input data sources. The Areas dataset contains data variables that have been standardised to ensure uniformity of the units and indices. The data variables in the Areas dataset are also documented to safeguard the transparency and reproducibility of the dataset.

The second step of the data processing approach is creating the Regions dataset, which contains the aggregated spatial data of the individual regions. The regions are defined by the spatial resolution reduction method applied and the selected countries of interest. The Regions dataset is, therefore, specific to the case study being investigated.

A Regions dataset is used in conjunction with additional techno-economic parameters and the calliope modelling framework to construct a power and heat optimisation model.

Heat sector modelling

The approach used to model the heat sector is depicted in Figure 3.2. The homogeneity of the input data used in the models is maximised by using the same temperature data to calculate both the COP factor of the ASHP and the space heating demand. As illustrated in Figure 3.2, the variation in ASHP COP and the space heating profiles are largely driven by air ambient temperature. As the days warm up during the depicted week the COP values increase and the heat demand for space heating decreases. The space heating demand in the colder months of the year makes up the largest share of the heat demand profile. The space heating per unit profiles is generated from a generic profile that provides normalised heat demand values for a range of ambient temperature values for different hours of the day. This generic profile was synthesised based on the performance of a heat pump installed in München, Germany [45]. The hot water demand profile is also generated by a generic profile that normalises hourly hot water heat demand profiles for nine typical days. Hotmaps have adjusted these generic profiles to account for the country dependent behaviours [46]. The end-use and sector-specific per unit heat demand profiles are multiplied by the total heat demand and their respective share factors obtained from hotmaps.

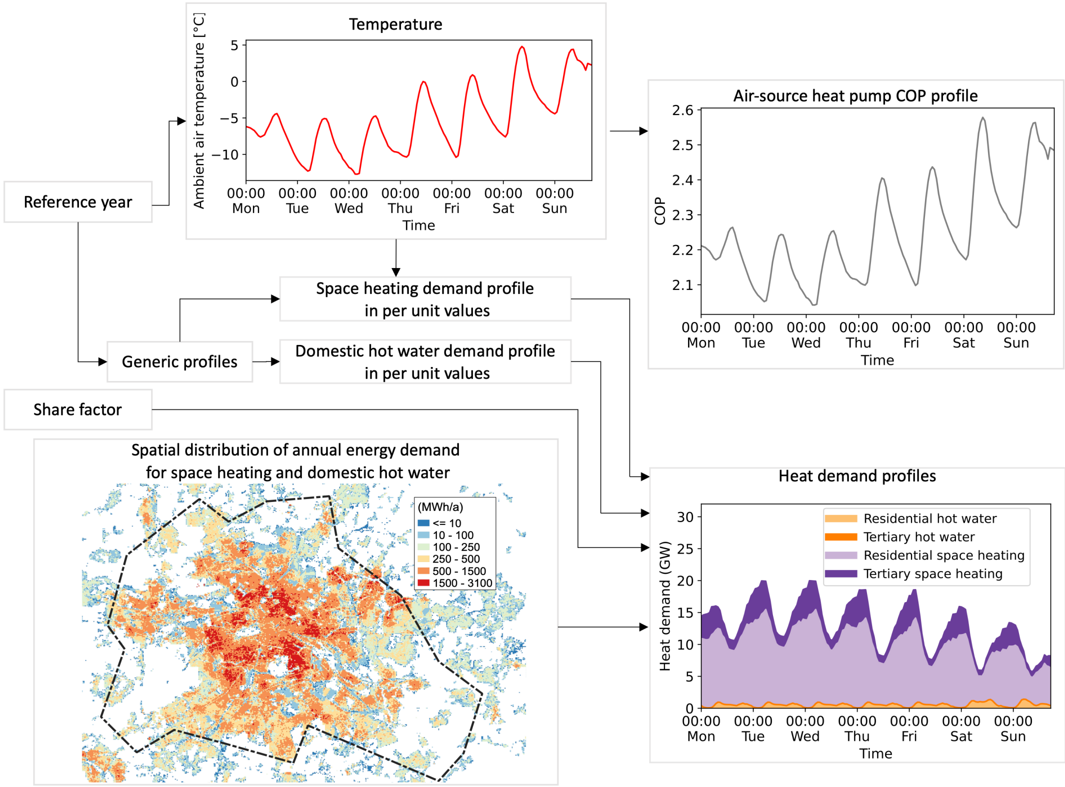


FIGURE 3.2: Visual description of the heat sector modelling method used to develop the power and heat optimisation models. The NUTS 2 area representing Berlin is used as an example for the illustration. The air-source heat pumps coefficient of performance profile and the heat demand profile represent a period of seven days in Febraury 2011.

3.2 Case studies

The models in the two papers differ in a spatial context in terms of geographic coverage and the spatial resolution reduction method applied. A summary of the case studies is given in Table 3.1. As the focus of **Paper A** is to investigate the effectiveness of the max-p regions in minimising spatial effects, several case studies with the different spatial scope were used. There are two spatial scope categories, country groups and individual countries.

Using a spatial scope that includes multiple countries, as is in-country groups, it is possible to compare the models built using the max-p-regions method with those built using the political regions method and the random-regions method. For individual countries, multiple regions were only built using the max-p-regions method and the random-regions method, as the political regions method will only generate a single region.

TABLE 3.1: The modelling frameworks, sectors and scenarios used to analyse the impact of different spatial resolution reduction methods on energy system models.

<table>
<tr><th rowspan="2">Paper</th><th rowspan="2">Modelling Framework</th><th rowspan="2">Sectors</th><th colspan="2">Scenarios and Case studies</th></tr>
<tr><th>Geographic coverage</th><th>Spatial scale reduction methods applied</th></tr>
<tr><td rowspan="2">Paper A</td><td rowspan="2">PyPSA</td><td rowspan="2">Power sector</td><td>Four country groups: Central West and the British Isles, South East, and North East</td><td>Max-p-regions, random-regions method and political regions method</td></tr>
<tr><td>Four individual countries: Germany, Spain, France and Italy</td><td>Max-p-regions and random-regions method</td></tr>
<tr><td rowspan="2">Paper C</td><td rowspan="2">Calliope</td><td rowspan="2">Power and heat sector</td><td>One country group: Germany, Norway, Denmark, Poland, France, Netherlands, Belgium, Austria, Switzerland, and the Czech Republic</td><td>Political regions method</td></tr>
<tr><td>One country: Germany</td><td>Max-p regions method and political regions method</td></tr>
</table>

Paper C focuses on demonstrating the variation in model results when different spatial resolution reduction methods are applied, and different spatial scopes are considered. In this demonstration, three spatial resolution reduction methods were used. The max-p-regions method and the political regions method (at the administrative NUTS 1 level) was used to build regions within the spatial scope of Germany. In contrast, the political regions method (at national jurisdiction NUTS 0 level) built regions within a larger spatial scope, including Germany and nine countries with a transmission connection with Germany.

3.3 Spatial resolution reduction methods

Spatial resolution reduction methods are used to build representative regions of the energy system being modelled. In **Paper A** and **Paper C**, there were a total of three spatial resolution reduction methods applied. These are the max-p-regions method, the random-regions method, the political regions. **Paper B** detailed the methods and materials needed to formulate regions using the max-p-regions method and the random-regions method.

The political regions method defines regions according to politically defined jurisdictions such as national jurisdiction (NUTS 0 level) or at sub-national administrative jurisdictions such as the NUTS 1 level for certain European countries. The concept of the max-p-regions method is to develop a method that is targeted at minimising the effects of spatial resolution reduction. The max-p-regions method aims to build regions that better characterise the areas they represent in terms of their energy-related spatial attributes. The spatial attributes selected are population, wind and solar resource potential, and pumped-hydro storage capacity. The population is used as a proxy for electricity or heat consumption.

The max-p-regions method and the random-regions method use the max-p-regions problem algorithm defined by She, Duque, and Ye [43]. A key benefit of the max-p-regions problem algorithm is that it contains a contiguous region constraint. This constraint ensures that all areas share at least one border with another area in the same regions. The pysal contiguity weights function [47] is used to determine the borders between areas. The max-p-regions method and the random-regions method uses the heuristic solution to solve the max-p-regions problem algorithm while specifying their respective spatial attribute. The spatial attribute used by the max-p-regions method is the energy-related attributes, while the random-regions method uses a set of random integers representing the areas.

In addition to the contiguous regions constraint, the MILP of the max-p-regions problem algorithm also requests a spatial attribute that defines the threshold value that all regions must attain. The threshold value used by the max-p-regions method and the random-regions method is the geometric area of the area. The use of geometric areas as the threshold value can make the max-p-regions problem algorithm unsolvable. The problem is unsolvable if the geometrical area of an island is lower than the set threshold value, as the pysal contiguity weights function islands does not assign borders with continental areas. This issue is avoided by manually assigning borders to islands with their closest continental areas.

Once regions have been defined, the spatial attributes are aggregated. As the areas in the regions differ in size, spatial attributes such as solar capacity factors, which are mean values, are aggregated using weights proportional to the areas' geometric area.

Results and discussions: Importance of spatial context

This section provides a summary of the result of the papers presented in this dissertation. Key insights gained from these papers are also discussed.

4.1 Results of papers

In Paper A, the effectiveness of the max-p-regions method in minimising two effects of spatial resolution reduction is investigated. The two effects of spatial resolution reduction are the increase in solar and wind expansion capacity and the decrease in transmission capacity expansion. Results show that model optimisation results from country groups, particularly the South-East Europe country group, that had regions built from using the max-p-regions method had on average higher transmission capacity expansion and lower solar and wind capacity expansion than optimisation results from country groups that were built using political regions method. Optimisation results of models of Germany and Spain built using the max-p-regions method showed that the max-p-regions method was an effective targeted approach of minimising the effects of spatial resolution reduction.

Paper B compliments Paper A in that it provides a clear and detailed description of the methodology of defining regions using the max-p-regions method and the random-regions method. Both novel methodologies use the NUTS 2 level areas of the European countries as the highest spatial resolution to apply the methods. The paper shows that the max-p-regions method and random-regions method uses a

MILP approach that is solved using the heuristic solution approach of the max-p-regions problem. The max-p-regions method uses energy-related attributes to define regions as a targeted approach of minimising the effects of spatial resolution reduction. Both spatial resolution reduction methods can generate as many regions within a spatial scope as there are NUTS 2 areas within geographical coverage of that spatial scope by varying area factor value used to calculate the minimum area threshold value of a region.

Paper C aims to present a novel data processing approach that incorporates spatial resolution reduction methods and maximises the use of existing web-hosted databases to build sector-coupled energy system optimisation models. The result of the paper shows how power and heat optimisation models of Germany with high penetration of solar and wind vary significantly depending on the spatial context of the model. The spatial context considered are the spatial scope and the spatial resolution reduction method applied to build the regions of the model. Significant variation in results was observed in the optimised generation capacity, transmission capacity and storage capacity expansion results.

4.2 Discussion of results

Below key takeaways from Paper A to Paper C are discussed.

4.2.1 Data processing in energy system modelling

Data processing is an important step in energy system optimisation modelling that, among other functions, organises the input data within a standardised structure that can be used to formulate an energy system model.

The aggregation or disaggregation of input data used to define the regions of a model is also a component of the data processing modelling step. As discussed and demonstrated in **Paper A** and **Paper C**, the spatial resolution reduction applied when building a model affects the optimisation results. Therefore the data processing approach used by an energy system model must be clearly described to allow for a proper understanding of the model results.

An objective of the novel data processing approach presented in **Paper C** is to maximise the use of existing dataset that are making documented pre-processed input data freely available. These datasets, such as the ENSPRESO database, offer various scenarios of capacity expansion areas in the NUTS 2 areas in Europe.

The scenarios define the eligible areas for wind and solar PV, which are multiplied by their respective power density to calculate the capacity expansion areas. Much of the work that generates these scenarios is the analysis of large georeferenced images or vectors used to create scenarios based on assumptions such as the distance of new installations from certain existing infrastructures. The analysis of such sizeable georeferenced information can be cumbersome and, in some cases, require sufficient computational capacity. The use of these scenarios thus avoids duplication of work, resulting in loss of valuable time that can be invested in other components of energy system modelling.

In the European energy modelling context, certain datasets, such as the ENSPRESO dataset, do not provide pre-processed data for Switzerland or Norway in their scenarios. This makes it more challenging to manage the homogeneity of the input data used in the model if complementary information specific to these two countries are needed to fill the gap.

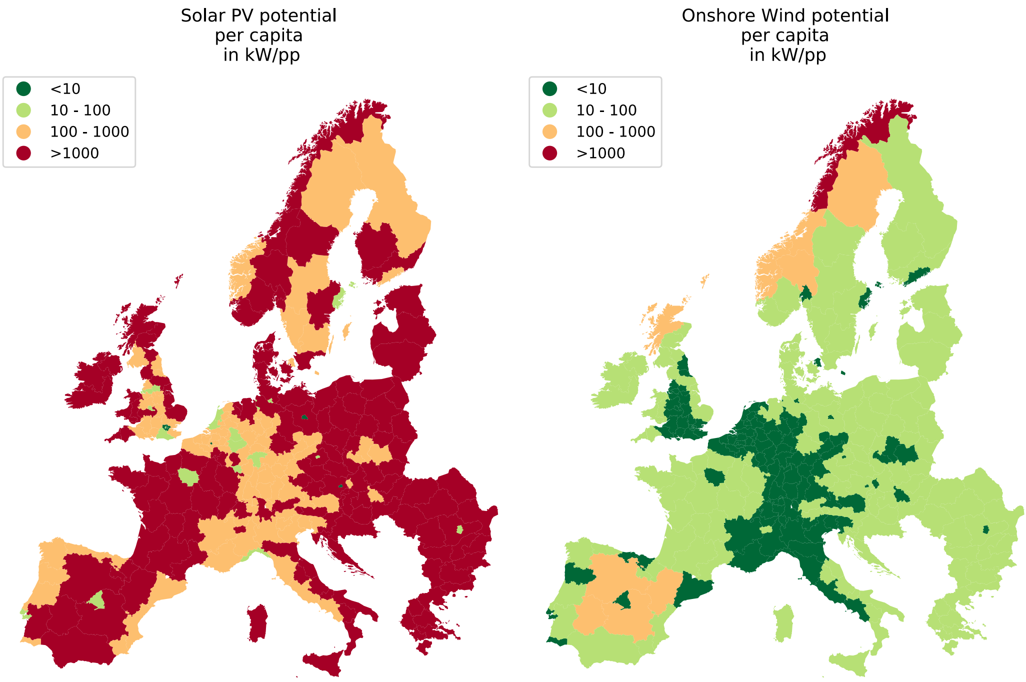


FIGURE 4.1: The figures show the ratio of potential PV and onshore wind capacity per capita in the NUTS 2 areas.

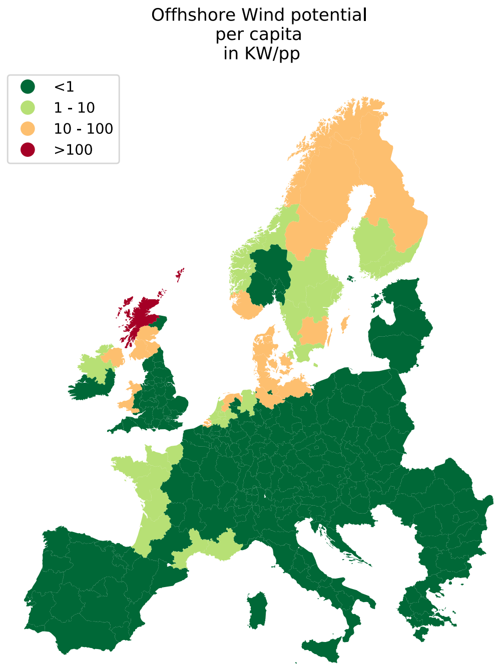


FIGURE 4.2: The figures show the ratio of offshore wind capacity per capita in the NUTS 2 areas.

The data processing approach presented can also be used to better comprehend the spatial and temporal relationship between the energy demand and VRE generation potential. Figure 4.1 shows the ratio of solar PV and onshore wind potential to the population in the NUTS 2 areas whereas 4.2 shows the same ratio per capita ratio for offshore wind. As described in **Paper C** the data on available area is primarily taken from the ENSPRESO dataset. The scenario in the dataset describing the potential calculation is described in Table 4.1. In **Paper A**, the power density of 12 MW/km^2 for solar PV is significantly lower than the power density of 170 MW/km^2 for solar PV in **Paper C** seen in Table 4.1. The power density value used in **Paper C** is closer to current literature values. The results from Enevoldsen and Jacobson [48] show that the power density values used in both **Paper A** and **Paper C** are more on the conservative side for onshore and offshore wind. The mean power densities of existing onshore and offshore wind farms in Europe indicated by Enevoldsen and Jacobson [48] are 19.8 MW/km^2 and 7.2 MW/km^2, respectively.

TABLE 4.1: Description of the assumptions used to determine the potential of solar PV, onshore wind, and offshore wind.

	Solar PV		Onshore wind	Offhore wind
	Rooftop PV	Utility		
Restriction	**Areas included**: Industrial and residential roofs and facades	**Areas included**: Natural and non-agricultural areas	**Areas excluded**: Water bodies, industry, forest, protected areas ..etc **Setback distance**: 400 m from settlement	**Areas excluded**: Depth >60m, shipping density > 5,000 vessels per year **Setback distance**: 12 NM from shore and 2 NM from shipping lanes, gas and oil pipelines, gas wells, submarine cables
Power density (MW/km^2)	170		5	5.36

4.2.2 The impact of spatial scale reduction methods that consider energy-related spatial attributes

The trackability of energy system optimisation models, especially for models with large spatial scopes, is limited by the spatio-temporal resolution of the model. The development of the max-p-regions method resulted from the gap in available spatial resolution methods that help reduce the computational demand of solving an energy system optimisation model while minimising the effects of spatial scale reduction.

The concept of the max-p-regions method proposed in Paper A and described in detail in Paper B is a regionalisation method that aims to generate a better representation of areas in energy system models. By taking an approach to regionalise areas into regions that better represent their areas, it is possible to mitigate spatial resolution reduction effects. For example, the inclusion of solar and wind resource potential as one of the three energy-related spatial resolution attributes used in the max-p-regions allows the regionalisation method to join areas with good solar and wind sites in similar regions. Good solar or wind sites may be regions with high resource wind or solar resource availability, high installation capacity potential or a combination of both. By forming regions of high solar and wind resource potential, the model optimisation maximises the use of these areas. These areas can then potentially become the energy-generating regions of the energy system. Another region may contain areas with higher population density, which could mean that these areas have less space for solar and wind energy production, resulting in higher transmission capacity needs to areas with higher energy production capacity. These regions can be identified for Germany, with the North of Germany having high wind resource potential while the south has a high demand for electricity and heat. As shown in Paper A, the max-p-regions method was efficient in minimising the spatial effects for the optimisation models simulated for Germany as it could identify these two types.

Therefore, the approach of the max-p-regions method builds energy systems that aim to build an energy system model that has a better understanding of the energy-related attributes of the areas it is simulating. By allowing energy system models to comprehend better the spatial aspects of the areas being modelled, these models will be more helpful in identifying areas that are better suited for specific energy-related activities.

4.3 Summarised contribution of the dissertation

This thesis contributes to improving energy modelling tools and the comprehension of how the spatial context of energy system models can impact optimisation results. Notably, this dissertation aims to establish methods that use spatial data to build models that are better targeted to the spatial context of the model and the problems being investigated. The application of regionalisation methods within the field of energy system optimisation modelling expands our understanding in terms of the possibilities in how energy systems can be spatially optimised to maximise the benefits that certain energy landscapes have to offer. These benefits could be good solar and wind resource potential or good electricity storage potential. The versatility of the data processing approach in defining regions using different spatial resolution methods allows for the construction of energy system models with a more targeted spatial specification. Although much of the work presented is technical, the results and advancement made in energy system modelling can have real-world effects. In the past, most large scale energy system models have used political boundaries to define regions. This approach can limit the energy-related policy discussions to these political boundaries. Using alternative methods of defining regions could present more pathways towards a sustainable energy system transition that is not limited to the political boundaries and more targeted to the specifications of the investigated areas. These alternative pathways can contribute and potentially improve energy-related policy discussions.

Conclusion

To improve the use of spatial data to define regions of energy system models, this dissertation:

- Presented a novel application of the max-p regions problem algorithm to generate a targeted method of reducing the spatial resolution of energy system models called the max-p regions method.

- Did a comparative analysis of how effective the max-p regions method was in minimising the effects of spatial resolution reduction.

- Presented a novel data-processing approach to build power and heat optimisation models for European countries with integrated spatial resolution reduction methods.

- Demonstrated the effects of spatial scale and choice of spatial resolution reduction methods on the results of power and heat optimisation models.

The comparative analysis conducted in **Paper A** shows that the max-p regions method can be used as a targeted method of minimising the effects of spatial resolution reduction. The max-p regions method can also be effective at minimising the effects of spatial resolution reduction when applied at national or at sub-continental geographical scale. The max-p regions are particularly effective when applied to case studies with significant distribution variations in spatial energy-related attributes such as energy generation potential from wind and solar.

From the results of case studies investigated in **Paper A** and **Paper C**, it can be concluded that the choice of spatial scale and spatial resolution reduction method can impact the results of power and heat optimisation models.

By developing novel methods of defining regions in energy system modelling and analysing the impacts of choice of spatial resolution reduction, it can be concluded that more importance must be placed on the use of spatial data in energy system modelling. The choice of regions can be tailored to the case study depending on spatial attributes' distribution to capture particular technology deployments' sensitivities better. The improved understanding of the role of spatial data and their impacts on energy system models will help reduce the uncertainties in model results used to inform decision-makers.

Research outlook

The work in this dissertation focuses on the techno-economic aspect of use of spatial data to inform energy system models. At some level there are also some socio-ecological aspects that are considered such as when defining the space available for deployment of solar and wind technologies. These socio-ecological aspects considered ensure that VREs are not deployed in areas that have been allocated towards nature conservation or areas that have existing infrastructures such as roads and airports.

More can be done to integrate energy-related socio-ecological aspects into energy system models. For example, Christ et al. [49] constrained the deployment of onshore wind to a defined balanced burden level in an energy system model of Germany to integrate socio-ecological factors of onshore wind. The methodology used to determine the balanced burden level considered the population density in the German districts and the available area for deployment of onshore wind. By including factors such as the burden level of certain technologies in the max-p regions method it may be possible to get a better representation of regions based on their suitability for deployment of particular technologies. Another socio-ecological aspect which could be improved and better integrated in energy system models is the determination of offshore area available for the deployment of offshore wind. For example the cost-effectiveness of offshore wind could be improved by the multi-use of offshore areas such as presented by Gusatu et al. [50]. An example of multi-use of offshore areas is the combination of offshore wind farms areas and fisheries activities which would require stakeholder engagement.

The proposed data processing method in **Paper C** offers an approach which is suited for the integration of new spatial attributes in datasets and spatial resolution reduction methods. These additional spatial attributes could also consider socio-ecological or even political aspects which can be quantified and which may affect energy system models.

Another area of potential future research in the area of use of spatial data is to improve energy system models by defining regions based on the spatial variability of power output of VREs. Research has shown that the strategical spatial deployment of wind power can offer cross-border balancing potential from from national to continental spatial scales [35, 36, 51]. Therefore, by considering spatial variability of VREs when defining regions of energy system models, the strategic spatial deployment could offer a more cost effective option in comparison to other flexibility options to manage the variability of VREs.

References

[1] Wilson, R., Gidden, M., Hare, B., and Pearson, P. "European Union 2030 emissions reduction target needs to be brought into line with the Paris Agreement 1.5oC limit". en. In: (Dezember 2020), p. 7.

[2] Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kobayashi, S., Kriegler, E., Mundaca, L., Séférian, R., Vilariño, M. V., Calvin, K., Emmerling, J., Fuss, S., Gillett, N., He, C., Hertwich, E., Höglund-Isaksson, L., Huppmann, D., Luderer, G., McCollum, D. L., Meinshausen, M., Millar, R., Popp, A., Purohit, P., Riahi, K., Ribes, A., Saunders, H., Schädel, C., Smith, P., Trutnevyte, E., Xiu, Y., Zhou, W., Zickfeld, K., Flato, G., Fuglestvedt, J., Mrabet, R., and Schaeffer, R. "Mitigation Pathways Compatible with 1.5 °C in the Context of Sustainable Development". en. In: (2018), p. 82.

[3] European Commission. Directorate General for Research and Innovation. *Electrification of the transport system: expert group report.* eng. LU: Publications Office, 2017.

[4] European Environment Agency. *Annual European Union greenhouse gas inventory 1990–2014 and inventory report 2016*. en. Tech. rep. type: dataset. Koninklijke Brill NV.

[5] Pfenninger, S., Hirth, L., Schlecht, I., Schmid, E., Wiese, F., Brown, T., Davis, C., Gidden, M., Heinrichs, H., Heuberger, C., Hilpert, S., Krien, U., Matke, C., Nebel, A., Morrison, R., Müller, B., Pleßmann, G., Reeg, M., Richstein, J. C., Shivakumar, A., Staffell, I., Tröndle, T., and Wingenbach, C. "Opening the black box of energy modelling: Strategies and lessons learned". en. In: *Energy Strategy Reviews* 19 (Jan. 2018), pp. 63–71.

[6] Pfenninger, S., Hawkes, A., and Keirstead, J. "Energy systems modeling for twenty-first century energy challenges". en. In: *Renewable and Sustainable Energy Reviews* 33 (May 2014), pp. 74–86.

[7] "The ERA5 global reanalysis". en. In: *Quarterly Journal of the Royal Meteorological Society* 146.730 (July 2020), pp. 1999–2049.

[8] Ryberg, D., Robinius, M., and Stolten, D. "Evaluating Land Eligibility Constraints of Renewable Energy Sources in Europe". en. In: *Energies* 11.5 (May 2018), p. 1246.

[9] Ryberg, D. S., Tulemat, Z., Stolten, D., and Robinius, M. "Uniformly constrained land eligibility for onshore European wind power". en. In: *Renewable Energy* 146 (Feb. 2020), pp. 921–931.

[10] Voulis, N., Warnier, M., and Brazier, F. M. T. "Understanding spatio-temporal electricity demand at different urban scales: A data-driven approach". en. In: *Applied Energy* 230 (Nov. 2018), pp. 1157–1171.

[11] Willis, H. and Northcote-Green, J. "Spatial electric load forecasting: A tutorial review". In: *Proceedings of the IEEE* 71.2 (Feb. 1983). Conference Name: Proceedings of the IEEE, pp. 232–253.

[12] Limei Zhou, Tao Wei, Mingtian Fan, and Zuping Zhang. "Study on the practical spatial load forecasting based on geographic information system". In: *2008 China International Conference on Electricity Distribution*. Guangzhou, China: IEEE, Dec. 2008, pp. 1–4.

[13] Pan, J. and Li, J. "Spatiotemporal Dynamics of Electricity Consumption in China". en. In: *Applied Spatial Analysis and Policy* 12.2 (June 2019), pp. 395–422.

[14] Mattsson, N., Verendel, V., Hedenus, F., and Reichenberg, L. "An autopilot for energy models – Automatic generation of renewable supply curves, hourly capacity factors and hourly synthetic electricity demand for arbitrary world regions". en. In: *Energy Strategy Reviews* 33 (Jan. 2021), p. 100606.

[15] Kanellopoulos, K., De Felice, M., Hidalgo Gonzalez, I., and Bocin, A. *JRC Open Power Plants Database (JRC-PPDB-OPEN)*. Version Number: 1.00 Type: dataset. Dec. 2019.

[16] Gimeno-Gutiérrez, M. and Lacal-Arántegui, R. "Assessment of the European potential for pumped hydropower energy storage based on two existing reservoirs". en. In: *Renewable Energy* 75 (Mar. 2015), pp. 856–868.

[17] "Electricity storage and renewables: Costs and markets to 2030". en. In: (2030), p. 132.

[18] Schlachtberger, D., Brown, T., Schramm, S., and Greiner, M. "The benefits of cooperation in a highly renewable European electricity network". In: *Energy* 134 (2017), pp. 469–481.

[19] Schaber, K., Steinke, F., Mühlich, P., and Hamacher, T. "Parametric study of variable renewable energy integration in Europe: Advantages and costs of transmission grid extensions". en. In: *Energy Policy* 42 (Mar. 2012), pp. 498–508.

[20] Medjroubi, W., Müller, U. P., Scharf, M., Matke, C., and Kleinhans, D. "Open Data in Power Grid Modelling: New Approaches Towards Transparent Grid Models". en. In: *Energy Reports* 3 (Nov. 2017), pp. 14–21.

[21] Novosel, T., Pukšec, T., Duić, N., and Domac, J. "Heat demand mapping and district heating assessment in data-pour areas". In: *Renewable and Sustainable Energy Reviews* 131 (2020), p. 109987.

[22] Möller, B., Wiechers, E., Persson, U., Grundahl, L., Lund, R. S., and Mathiesen, B. V. "Heat Roadmap Europe: Towards EU-Wide, local heat supply strategies". In: *Energy* 177 (2019), pp. 554–564.

[23] Martínez-Gordón, R., Morales-España, G., Sijm, J., and Faaij, A. "A review of the role of spatial resolution in energy systems modelling: Lessons learned and applicability to the North Sea region". en. In: *Renewable and Sustainable Energy Reviews* 141 (May 2021), p. 110857.

[24] Hörsch, J., Hofmann, F., Schlachtberger, D., and Brown, T. "PyPSA-Eur: An Open Optimisation Model of the European Transmission System". en. In: *Energy Strategy Reviews* 22 (Nov. 2018). arXiv: 1806.01613, pp. 207–215.

[25] Pfenninger, S. and Pickering, B. "Calliope: a multi-scale energy systems modelling framework". In: *Journal of Open Source Software* 3.29 (2018), p. 825.

[26] Hilpert, S., Kaldemeyer, C., Krien, U., Günther, S., Wingenbach, C., and Plessmann, G. "The Open Energy Modelling Framework (oemof) - A new approach to facilitate open science in energy system modelling". In: *Energy Strategy Reviews* 22 (2018), pp. 16–25.

[27] Tröndle, T. *euro-calliope: pre-built models*. en. Version Number: 1.0 Type: dataset. July 2020.

[28] Hörsch, J. and Brown, T. "The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios". In: *2017 14th International Conference on the European Energy Market (EEM)*. ISSN: 2165-4093. June 2017, pp. 1–7.

[29] Shawhan, D. L., Taber, J. T., Shi, D., Zimmerman, R. D., Yan, J., Marquet, C. M., Qi, Y., Mao, B., Schuler, R. E., Schulze, W. D., and Tylavsky, D. "Does a detailed model of the electricity grid matter? Estimating the impacts of the Regional Greenhouse Gas Initiative". In: *Resource and Energy Economics* 36.1 (2014), pp. 191–207.

[30] Frew, B. A. and Jacobson, M. Z. "Temporal and spatial tradeoffs in power system modeling with assumptions about storage: An application of the POWER model". en. In: *Energy* 117 (Dec. 2016), pp. 198–213.

[31] Frysztacki, M. M., Hörsch, J., Hagenmeyer, V., and Brown, T. "The strong effect of network resolution on electricity system models with high shares of wind and solar". en. In: *Applied Energy* 291 (June 2021), p. 116726.

[32] Krishnan, V. and Cole, W. "Evaluating the value of high spatial resolution in national capacity expansion models using ReEDS". In: *2016 IEEE Power and Energy Society General Meeting (PESGM)*. Boston, MA, USA: IEEE, July 2016, pp. 1–5.

[33] Simoes, S., Zeyringer, M., Mayr, D., Huld, T., Nijs, W., and Schmidt, J. "Impact of different levels of geographical disaggregation of wind and PV electricity generation in large energy system models: A case study for Austria". en. In: *Renewable Energy* 105 (May 2017), pp. 183–198.

[34] Tröndle, T., Lilliestam, J., Marelli, S., and Pfenninger, S. "Trade-Offs between Geographic Scale, Cost, and Infrastructure Requirements for Fully Renewable Electricity in Europe". en. In: *Joule* 4.9 (Sept. 2020), pp. 1929–1948.

[35] Grams, C. M., Beerli, R., Pfenninger, S., Staffell, I., and Wernli, H. "Balancing Europe's wind-power output through spatial deployment informed by weather regimes". en. In: *Nature Climate Change* 7.8 (Aug. 2017), pp. 557–562.

[36] Santos-Alamillos, F., Pozo-Vázquez, D., Ruiz-Arias, J., Lara-Fanego, V., and Tovar-Pescador, J. "A methodology for evaluating the spatial variability of wind energy resources: Application to assess the potential contribution of wind energy to baseload power". en. In: *Renewable Energy* 69 (Sept. 2014), pp. 147–156.

[37] Tröndle, T., Pfenninger, S., and Lilliestam, J. "Home-made or imported: On the possibility for renewable electricity autarky on all scales in Europe". In: *Energy Strategy Reviews* 26 (2019), p. 100388.

[38] Müller, U. P., Schachler, B., Scharf, M., Bunke, W.-D., Günther, S., Bartels, J., and Pleßmann, G. "Integrated Techno-Economic Power System Planning of Transmission and Distribution Grids". en. In: *Energies* 12.11 (May 2019), p. 2091.

[39] Temraz, H., Salama, M., and Quintana, V. "Application of partitioning techniques for decomposing large-scale electric power networks". In: *International Journal of Electrical Power & Energy Systems* 16.5 (1994), pp. 301–309.

[40] Anderski, T., Surmann, Y., Stemmer, S., Grisey, N., Momot, E., Leger, A.-C., Betraoui, B., and Van Roy, P. *D2.2 European cluster model of the Pan European transmission grid*. Tech. rep. 2015.

[41] Duque, J., Dev, B., Betancourt, A., and Franco, J. *ClusterPy: Library of spatially constrained clustering algorithms, Version 0.9.9*. Jan. 2011.

[42] Siala, K. and Mahfouz, M. Y. "Impact of the choice of regions on energy system models". en. In: *Energy Strategy Reviews* 25 (Aug. 2019), pp. 75–85.

[43] She, B., Duque, J. C., and Ye, X. "The Network-Max-P-Regions model". en. In: *International Journal of Geographical Information Science* 31.5 (May 2017), pp. 962–981.

[44] Getman, D., Lopez, A., Mai, T., and Dyson, M. *Methodology for Clustering High-Resolution Spatiotemporal Solar Resource Data*. en. Tech. rep. NREL/TP-6A20–63148, 1225896. Sept. 2015, NREL/TP–6A20–63148, 1225896.

[45] GmbH, S. M. *Lastprofil Wärmepumpe*. de.

[46] Pezzutto, P. S., Zambotti, S., Croce, S., Zambelli, P., Scaramuzzino, C., Pascuas, R. P., Zubaryeva, A., Exner, D., Müller, A., Klingler, A.-L., Kühnbach, M., Manz, P., Marwitz, S., Steinbach, J., Popovski, E., Kranzl, R. L., and Fritz, S. "D2.3 WP2 Report – Open Data Set for the EU28". en. In: (2018 2018), p. 158.

[47] Rey, S. J. and Anselin, L. "PySAL: A Python Library of Spatial Analytical Methods". en. In: *Handbook of Applied Spatial Analysis*. Ed. by M. M. Fischer and A. Getis. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 175–193.

[48] Enevoldsen, P. and Jacobson, M. Z. "Data investigation of installed and output power densities of onshore and offshore wind turbines worldwide". In: *Energy for Sustainable Development* 60 (2021), pp. 40–51.

[49] Christ, M., Soethe, M., Degel, M., and Wingenbach, C. "Wind Energy Scenarios for the Simulation of the German Power System Until 2050: The Effect of Social and Ecological Factors". In: *Advances and New Trends in Environmental Informatics*. Ed. by V. Wohlgemuth, F. Fuchs-Kittowski, and J. Wittmann. Series Title: Progress in IS. Cham: Springer International Publishing, 2017, pp. 167–180.

[50] Gusatu, L. F., Yamu, C., Zuidema, C., and Faaij, A. "A Spatial Analysis of the Potentials for Offshore Wind Farm Locations in the North Sea Region: Challenges and Opportunities". en. In: *ISPRS International Journal of Geo-Information* 9.2 (Feb. 2020). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 96.

[51] Wohland, J., Brayshaw, D. J., and Pfenninger, S. "Mitigating a century of European renewable variability with transmission and informed siting". In: *Environmental Research Letters* (May 2021).

Part II: Publications

Minimising the effects of spatial scale reduction on power system models

Paper A

Christian Etienne Fleischer [a] [a] Department of Energy and Environmental Management, Europa-Universität Flensburg, Flensburg, Germany

Abstract

The spatial resolution of energy system optimisation models (ESOMs) is often compromised for computational performance. Reducing the spatial resolution impacts the least-cost solutions when optimising the generation capacity and transmission capacity of the ESOMs with high penetration of variable renewable energy sources. Previous studies show that the two main effects of spatial resolution reduction are the increase in solar and wind expansion capacity and the decrease in transmission capacity expansion. This paper introduces a targeted method of defining regions during spatial scale reduction by using the max-p-regions clustering algorithm to aggregate similar areas into regions. The attributes used to determine the similarity between areas in the max-p-regions method are population; wind and solar resource potential; and pumped-hydro storage capacity. Two alternative spatial resolution reduction methods were used to compare how the effects of spatial resolution impacted the optimisation results of the ESOMs. Evaluation results for three country groups showed that using the max-p-regions method to define regions caused the two effects of spatial resolution reduction to be generally lower than using national jurisdictions. For the case studies Germany and Spain, the results showed that the max-p-regions method identified sets of regions, which were less impacted by the two effects of spatial resolution reduction.

Keywords: Spatial analysis, Transmission expansion, Energy system modelling, Renewables

Paper published in Energy Strategy Reviews Volume 32 (2020)
https://doi.org/10.1016/j.esr.2020.100563

A.1 Introduction and state of the art

There is high confidence that electricity will need to take up a larger share of the final energy demand by 2050 to avoid a 1.5°C warming of the earth. The electricity will also have to be generated primarily by renewable energy technologies [1]. One critical consideration for minimising the costs associated with the energy system transition is the placement of variable renewable energy resources (VRE), including solar and wind energy technologies. Because local weather conditions dictate their performance, VRE can create bottlenecks when transmitting the electricity generated to regions where it is to be consumed or stored. Energy system optimisation models (ESOMs) are used to ensure the efficient matching of supply and demand of different VREs at a relevant spatio-temporal scale. Importantly, an increase in the spatiotemporal resolution of ESOMs leads to an increase in the computational demand for solving the investment optimisation of these models [2]. [3] identified that balancing the spatial and temporal resolution with data availability and computational tractability as one of the main challenges facing large bottom-up optimisation models.

Several methods are used by the modelling community to maintain computational tractability of ESOMs. Temporally, models such as the JRC-EU-TIMES model [4] created time slices that represent intra-annual variations in demand and supply of the optimisation problem. This method is common for modelling tools that have multi-year modelling horizons, including LEAP [5] and OSeMOSYS [6]. Other methods, for instance, EnergyPlan, did not use time-slices but limited the time horizon of the model to one year [7]. Some of these models achieved tractability by compromising on the spatial resolution of the model. [2, 8, 9] discuss how the choice in spatial resolution impact the results of large-scale energy system models with high penetrations of renewables. [8] and [9] focuses on the United States (US) power system, while [2] concentrates on the European power system. In all three studies, decreasing the spatial resolutions of the ESOMs produces higher, sub-optimal, total system costs. The oversizing of wind and solar capacity contributed to an increase in total cost. [2] found that the increase in spatial detail increases the transmission line cost, as there is an increase in intra-country bottlenecks that separate load centres from areas with good wind and solar sites. [8] compared the impacts of temporal and spatial resolution reduction on the total optimised investment cost. The comparison showed that a reduction in spatial resolution has a greater impact on the total minimised system cost than the impact of temporal resolution reduction. At a smaller geographical scale, [10] analysed the impact of different levels of geographical disaggregation of wind and solar on an energy system model of Austria. The study argued that the spatial resolution representation of renewables has a greater effect on countries where differences in performance of renewables are more pronounced [10].

With the increased availability of higher spatial resolution data, new methods are emerging and changing how ESOMs regions are defined. [8] and [10] identified and created regions that are location and technology-specific—differing in their availability data and cost parameters attributed. Another method [2], used georeferenced electrical grid data and clustering algorithms to define regions. Specifically, they used k-means clustering and power system attributes at electrical nodes to define the spatial resolution of the model [2]. The benefit this method is that it maintains the major transmission corridors, although it requires access to an extensive, in this case, European-wide electrical system dataset. Similarly, two other approaches, [11] and [12], used clustering algorithms to define regions but instead applied it to geographical attributes that can reflect wind and solar resource potential. [11] identified similar regions using several characteristics that are linked to electricity consumption and generation potentials of regions. This approach aimed at identifying regions most likely to exchange electricity and thus potentially distinguishing key transmission capacity needs. The k-means clustering algorithm in [11] used energy-related attributes to identify similar regions. These energy-related attributes include wind speed, solar irradiation, population, hydro-power plants capacity and agricultural areas and natural grasslands. Also, [11] amended the clusters created for European countries after consultation with Transmission System Operators (TSOs) that advised on other criteria's that required the clusters to be adjusted. Although [11] and [12] both created regions based on the similarities between areas within the same regions, there are still several key distinctions between their clustering approaches. Firstly, [11] used several attributes to identify one set of regions, while [12] used only one spatial attribute as input to the clustering algorithm. Secondly, [12] used both the max-p-regions and the k-means algorithm in the clustering process. The max-p-regions algorithm appears to be less commonly used in the energy modelling field. However, it has one key advantage over to the k-means algorithm, in that it guarantees spatial contiguity of the regions. All areas in a spatially contiguous region share at least one border with another area in the same region. Results from [12] showed that the use of national jurisdictions to define regions, leads to higher generation capacity needs, compared using similarities between areas based on either solar, wind or energy demand spatial attributes.

By expanding on the methods described in [11] and [12], this paper presents a targeted approach to define regions to reduce the spatial resolution of energy systems. The objective of the approach, referred to as the max-p-regions method, is to create regions that are similar in their electricity consumption, generation and storage capabilities during spatial resolution reduction. The max-p-regions method is compared with two other spatial resolution reduction methods based on how it is affected by the recognised effects of spatial resolution reduction on ESOMs. The two effects of spatial resolution reduction assessed are the increase in solar and wind capacities [2, 8, 10] and the reduction in transmission capacity expansion [2, 8].

The introduction is followed by Section A.2, which describes the max-p-regions method and the alternative methods used to define regions during spatial resolution reduction. The first spatial resolution reduction method used is a non-targeted approach which generates random sets of regions. The second method defines regions based on the national jurisdiction of areas. The ESOM structure and the input data used to evaluate the effects of spatial resolution reduction is elaborated upon in Section A.3. In Section A.4, the different spatial resolution reduction methods are evaluated on how they impact the effects of spatial resolution reduction of ESOMs. The evaluation is carried out on four individual European countries and three groups of European countries. The results of the evaluation are summarised and discussed in Section A.5.

A.2 Spatial resolution reduction methods

A.2.1 Max-p-regions method

The max-p-regions algorithm, presented in [13], clusters areas into an unknown number of regions by minimising an objective function of a mixed integer programming (MIP) model. The objective function of the algorithm aims to minimise dissimilarities between areas in the same region while also maximising the number of regions generated. The user selects the attributes of the areas used to define their heterogeneity. The MIP model gives precedence of maximising the number of regions over reducing total heterogeneity. The objective function is subject to multiple constraints, including the minimum threshold constraint, which ensures that all regions surpass a threshold of a spatially extensive attribute. The max-p-regions problem algorithm also uses an initial seed value to select which areas are selected to begin the clustering process, thus varying this seed value can alter the regions created.

As described in more detail in Section A.3.2, the input data is first aggregated to Nomenclature of territorial units for statistics level 2 (NUTS 2) areas. Depending on the European country, a NUTS 2 area jurisdiction can include the whole country or only a sub-national economic territory. The max-p-regions function in the open-source pysal python package [14] is used to apply the max-p-regions problem algorithm on these NUTS 2 areas. When using the max-p-regions function, three energy-related attributes of the regions determine the dissimilarities between areas and regions. These attributes are population; wind and solar resource potential; and pumped-hydro storage capacity. The population is the proxy used for energy demand. Existing pumped hydro storage capacity is used to measure a regions ability to store large amounts of energy. Solar and wind energy potential is the indicator of the amount of solar and wind energy a region can generate. Appendix A.A illustrates the spatial distribution of these three attributes across NUTS 2 areas of country groups. Appendix A.B illustrates the same attributes of NUTS 2 areas of Germany, Spain, France and Italy.

Two sets of matrices are used in the pysal max-p function to capture spatial contiguity between areas. The first matrix contains a list of neighbouring areas for all areas, whereas the second matrix sets a weighting value to each area in the lists. Two areas are neighbours if they share at least one vertex. The spatially extensive attribute used in the max-p-regions function to define the minimum threshold value that each region should surpass is the area of the NUTS 2 areas. The median area of the NUTS 2 areas is multiplied with a factor to determine the minimum threshold value. By adjusting this factor, it is possible to either increase or decrease the number of regions that will be created by the max-p-regions function.

For countries with islands, after a certain minimum threshold area value, the objective function of the MIP model is not solvable. This is because there are NUTS 2 areas that are islands and they do not share vertexes with any of the continental NUTS 2 areas. This issue was avoided by manually grouping islands with the nearest continental area. Appendix A.C illustrates the grouped islands and continental areas. The parameters of the areas in the same group or region were aggregated to obtain a single representative set of parameters. Parameters that are spatial average values, such as per unit maximum power output of wind turbine generators (WTGs), photovoltaic systems (PV) and run-of-river, were assigned weights that are proportional to the area they are representing during aggregation. All other parameters of the NUTS areas were summed together during aggregation. Regions created using the max-p-regions method is hereafter referred to as max-p-regions.

A.2.2 The random-regions method

The random-regions clustering method defines regions using a non-targeted approach while maintaining spatial contiguity of the region. It uses the same pysal max-p function used by the max-p-regions method. The attributes given to the function to define heterogeneity between regions are random values between 0 and 1. A predefined initial seed parameter and a python package called numpy [15] generated the random values. The random-regions method can create potentially all possible clusters sets of regions, within the constraints of the max-p function. Similar to the max-p-regions, the number of regions generated by the random-regions depends on the minimum threshold value. Regions created using the random-regions method is hereafter referred to as random regions.

A.2.3 The random-regions method

The political-regions method aggregates spatial data of NUTS 2 areas based on the national jurisdiction it belongs to. Energy system models commonly use this spatial aggregation method when the model encompasses more than one country. Regions created using the political-regions method is hereafter referred to as political regions. Fig. A.1 gives an illustration of how the different spatial resolution methods define the regions.

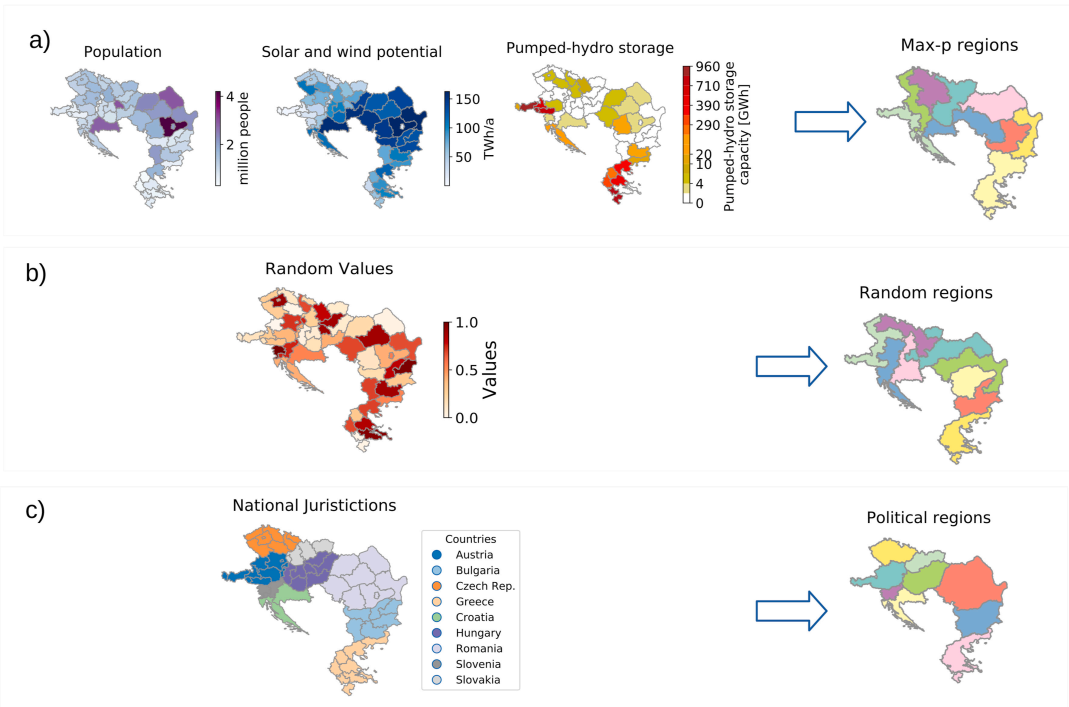


FIGURE A.1: A visual example of how the a) max-p-regions, b) random-regions and c) political-regions methods are used to create different sets of regions from the areas of some countries located in the South-East of Europe.

A.3 Power system optimisation model and input data

A.3.1 Model structure

Power system optimisation models are the ESOMs used to evaluate the spatial resolution reduction methods. The possible direction of energy flows between components in the same region are displayed in Fig. A.2. The energy carriers in the model are electricity and gas. Although both energy carriers can travel between regions, electricity transmission is limited to the capacity of the lines between regions. The power system optimisation models are modelled using the PyPSA framework [16]. The objective function of the linear optimisation used to minimise the investment and dispatch cost of the power system, for a period of one year and a time resolution of 3 h, is given by

$$\min_{G_{n,s},F_l,g_{n,s,t}} [\sum_{n,s} c_{n,s} \cdot G_{n,s} + \sum_{l} c_l \cdot F_l + \sum_{n,s,t} o_{n,s} \cdot g_{n,s,t}] \tag{A.1}$$

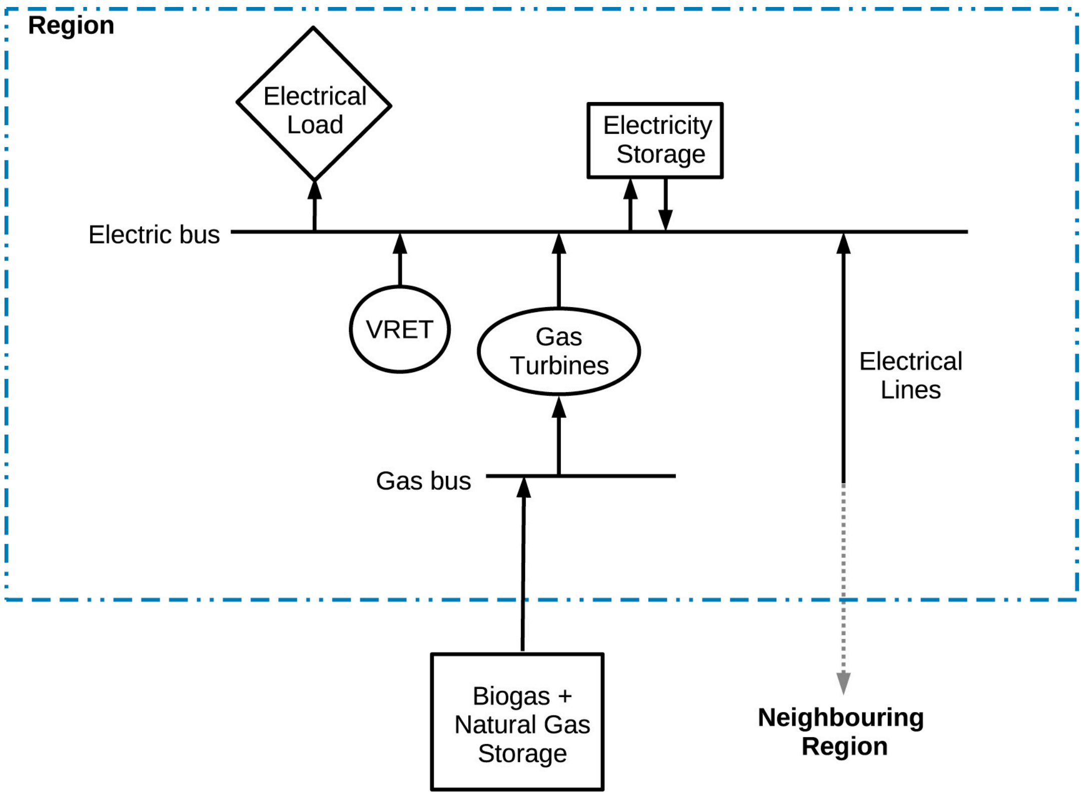


FIGURE A.2: Structure of the power system optimisation models used to capture the effects of spatial resolution reduction. The arrows show the direction of possible energy flow between components.

The investment costs include the annualised costs $c_{n,s}$ for the installed capacity of the generation and storage technologies $G_{n,s}$ and the annualised investment costs c_l of increasing the transmission lines capacity F_l. The operational costs are the variable costs $o_{n,s}$ occurred, at each time interval t, for energy unit dispatched for generation or storage $g_{n,s,t}$. Constraints ensure the balance of power at each node and instance; the dispatch of energy is not greater than the installed capacity; the weather-dependent technologies dispatch per the limits of the weather conditions. [16] gives a detailed description of the constraints applied to the objective function of the power system optimisation models. The weather-dependent technologies in this model are onshore and offshore wind turbines, PV systems and run-of-river plants. The optimisation assumes: perfect competition; perfect foresight; and the energy demand is inelastic. The installed capacity of certain technologies can be increased at a cost. These include onshore and offshore wind, PV systems, gas turbines, battery packs and hydrogen storage. Space availability limits the increase in capacity of PV systems and onshore and offshore wind. The methodology to calculate the maximum allowable increase in installed capacity is described in Section A.3.2.

A.3.2 Input data

This section will elaborate on the input data used in the model. All input data have been aggregated to represent the NUTS 2 areas. The NUTS 2016 geometry (1:10 Million scale) and standardised ID codes of NUTS 2 areas have been extracted from the Eurostat database [17].

The European Network of Transmission System Operators (ENTSO-E) transparency platform [18] provides hourly power profiles for European countries. These country-specific profiles are used in combination with rasterized population data [19] to obtain profiles for individual NUTS 2 areas. The population data is used to obtain fraction values of population distribution, for each area, which are assumed to align to load distribution. These fraction values are factored to the hourly profiles to obtain profiles for each area. The basis year chosen for the load profile is 2010. The aggregated installed capacity of hydropower (run-of-river and reservoir) and gas power plants within each area were extracted from the Global Power Plant Database (GPPD) [20]. The weather-dependent power output profiles of wind turbines and PV systems are given to the model as normalised maximum power output profiles extracted from renewables.ninja platform [21]. The maximum potential yield profiles of onshore wind and PV systems are provided at NUTS 2 level whereas for offshore wind they are only country-specific. The hydropower plant power dispatch is also limited by daily inflow, which is given for each country in [22]. The inflow data for the year 2010 was used to model the allowable hydro plant maximum dispatch profiles.

During the power system optimisation, wind and PV systems can increase their installed capacity at a cost. To calculate the maximum allowable capacity increase, the sum of the areas which are eligible for the installation of these VREs have been estimated. This is undertaken using the Geospatial Land Availability for Energy Systems framework (GLAES)[23]. GLAES eliminates non-eligible areas from the total area being considered to identify the remaining eligible areas. Non-eligible land areas are areas that have defined utility purposes, such as settlement proximity, protected parks and power lines, that disqualifies them from the installation of VREs. Distance parameters are used to determine the buffer zones that are also considered as non-eligible areas. The utility purposes considered and set distance parameters used for this study to identify eligible land areas for onshore wind or PV is defined in the med scenario from [23]. During optimisation of the power system model investment, the identified eligible land areas can only be used once by either onshore wind or PV systems. A density of 4 MW/km^2 and 12 MW/km^2 is assumed for onshore wind farms and PV farms respectively. The GLAES tool is also used to identify the eligible sea area for offshore wind within exclusive economic zones (EEZ) of the European countries. The criteria for non-eligible areas are: areas listed in the World Database on Protected Areas (May 2019 version) [23]; areas with depths greater than 60 m; areas wherein 2017 ships were recorded to have spent in average more than 1 h in a square km per month; areas within 12 nautical miles of the coast; areas within 1 nautical mile of gas and oil pipelines [24]. Only 30% of the remaining eligible area is considered available for wind farms, to account for other marine time uses within the EEZ. An offshore wind farm density of 5.36 MW/km^2 is used to calculate the total maximum allowable wind farm installed capacity. This capacity value is shared proportionally amongst areas with a coastal area that is at least a 5th percentile of the total coastal area of the country.

In the model, electricity can be generated by combusting biogas or natural gas in gas-fired power plants identified in the GPPD. Both resources are limited in quantity with a respective annual budget. The annual budget of natural gas is derived from an allowable net C02 emission budget of the power sector model. The total natural gas emission budget is 5% of the CO_2 equivalent emissions of the power sector in 1990 of EU28 plus Norway and Switzerland. The sum in CO_2 equivalent emissions of the power sector in 1990 for these countries is taken to be 1360 megatonnes of CO_2 ($MtCO_2$) [25] and the natural gas emissions factor as 0.19 $tCO_2/MWhth$. The total budget of natural gas is then calculated to be 358 TWhth/a. The total annual budget of biogas allocated for these European countries is limited by space availability and is assumed to be a total of 63 TWhel [26]. Both the natural gas and biogas budget is

TABLE A.1: The financial and technical parameters of energy storage technologies used in the optimisation of power system models.

Technology Type	Investment Cost			Operation and Maintenance Cost		Efficiency	Lifetime	Maximum hours
	Charging unit	Discharging unit	Storage unit	Fixed	Variable			
	€/kW	€/kW	€/kW	€/kW	€/kWh		years	hours
PHS	–	–	–	11	0.05	0.76	–	6
Battery	0	80	855	1.6	0	0.95	20	6
Hydrogen	645	727	450	10.8	0.3	0.41	30	128

then proportionally distributed to each country based on their population.

The electrical grid is modelled as a high voltage alternative current (HVAC) network. It is assumed that the HVAC network is operating at 380 kV. The electric buses have the same geographical coordinates of the centroid of the regions, and the bordering regions are interconnected using two HVAC lines. The lines have a resistance of 0.059 Ω/km and installed current-carrying capacity of 960 A. In addition to the HVAC network, there are high voltage direct current links (HVDC) that link certain NUTS 2 areas together. It is assumed that these DC links operate at an efficiency of 96%. In Appendix A.E, both the existing and planned DC links have been summarised from various references. It is also assumed that the DC links and the HVAC lines installed carrying capacity can be extended at the cost of 600 €/MWkm.

The rechargeable energy storage technologies in the model are pumped hydro storage (PHS), batteries and hydrogen. The location and storage capacity of PHS plants that are either labelled as pure PHS or reservoir based hydropower plants from the JRC hydropower plants database [27] was used in the model. In contrast to PHS, the storage capacities of battery storage and hydrogen can be increased at a cost. The financial and technical details of the storage technologies are provided in Table A.1 taken from [28] for the year 2020. The capital cost for increasing the installed capacity of storages are annualised sum of investment cost plus the fixed operation and maintenance cost. The financial and technical parameters for generation technologies assumed are projected values for 2020 taken from [29] and are given in Table A.2 .

TABLE A.2: The financial and technical parameters of the energy generation technologies used in the optimisation of power system model. The connections between applying the spatial resolution reductions methods and creating the optimisation models is illustrated in Appendix A.D.

Technology Type	Investment Cost	Operation and Maintenance Cost		Efficiency	Lifetime
		Fixed	Variable		
	€/kW	€/kW/a	€/MWh		years
Hydro	NA	0.074	15	0.90	80
Gas	495	6.5	49	0.46	25
Solar PV	830	8	0	1	20
Wind onshore	1120	14	1.5	1	27
Wind Offshore	2130	40.6	3	1	27

TABLE A.3: Groups of European countries used as case studies.

No.	Name	Countries	Number of NUTS 2 areas	Mean area per NUTS 2 areas (km^2)
1	Central West and British Isles	Great Britain, Ireland, Belgium, Spain, France, Portugal, Switzerland, Italy, Luxembourg	123	14,735
2	South East	Czech Republic, Slovakia, Greece, Hungary, Bulgaria, Romania, Croatia, Austria	53	38,215
3	North East	Denmark, Germany, Poland, Estonia, Latvia, Lithuania, Finland, Sweden, Norway, Netherlands	94	8755

A.4 Evaluation

A.4.1 Country groups

In this section, the three spatial resolution reductions methods, described in Section A.2, are applied to the NUTS 2 areas of three groups of European countries. These country groups are described in A.3.

The political-regions method defines the same number of regions as there are countries in the group. To standardise the spatial resolution reduction methods, the selected minimum threshold values, created the same number of regions as the number of countries in the country groups. As mentioned in Section A.2, regions defined by the max-p-regions method and random-regions method can vary depending on the initial seed value used. Thus, the initial seed values were varied to create three sets of max-p-regions and three sets of random regions to capture this potential variation. With the addition of one set of political regions, there were a total of seven sets of regions created per country group. Each set of region formulated one power system model, which was then optimised.

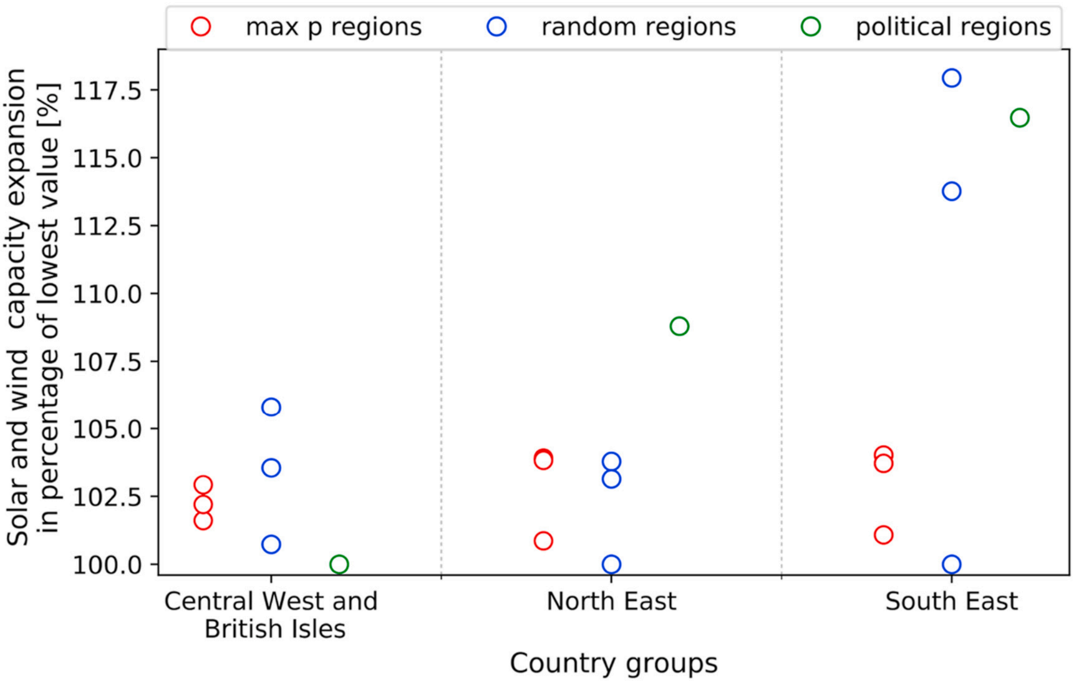


FIGURE A.3: The sums of the solar and wind capacity expansions obtained from the optimisation results of the regions sets, given in percentage of the lowest value in the same country group.

Fig. A.3 plots the sum of solar and wind generation capacity expansion from the optimisation results of each power system model. The sum values are categorised according to the spatial resolution reduction method used and the respective country group. The solar and wind capacity expansion values are given in percentage of the lowest value obtained from the results of the optimisation of the power system models in the same country group [1]. The lowest value is the optimisation model that was relatively least impacted by the effect of spatial resolution reduction on solar and wind expansion capacity. The transmission capacity expansion values are given in percentage of the highest value obtained for the spatial resolution reduction methods within the same country group. The highest value is the optimisation model that was relatively least impacted by the effect of spatial resolution reduction on transmission line expansion capacity.

Fig. A.3 shows that in the case of the South East and North East country groups, applying the max-p-regions method resulted in the solar and wind capacity expansion values being lower than the values from the political region. Particularly the results from the South East region show the political regions value being 13% greater than the lowest value obtained from using the max-p-regions method. In the case of the results from the Central West and the British Isles country group, the max-p-regions method did not identify a set of regions which generate a lower solar and wind expansion value than that of the political-regions method. The greatest difference between the values obtained from this country group, including those obtained from the random regions, does not exceed 7%. This difference is lower than that of the other country groups, potentially indicating less flexibility in the formation of regions that could generate lower solar and wind expansion values. The results from the figure also show that the values from random regions can range from the lowest and highest solar and wind expansion values. Relatively, the difference between the values generated by max-p-regions is smaller compare to the difference between values generated by the random regions.

The transmission line capacity expansion values from the optimisation results are plotted in Fig. A.4. For all country groups, there is a max-p-region transmission line capacity expansion value which is greater than the value from the political regions. Particularly in the South East country group, the political region value is less than half the value of the highest value obtained from the max-p-regions method. In general, the results from the country groups indicate that the use of max-p-regions can define regions of a power system optimisation model that is less affected by the effects spatial resolution reduction in comparison to the use of political regions method. Particularly in the case of the South East country group, the results indicate that the max-p-regions are on average less affected by the effects of spatial resolution reduction than the random regions. The exercise shows that the range of values obtained from using the random-regions method is on most occasions greater than the max-p-regions method and encompasses the values obtained from the political-regions method.

A.4.2 Individual countries

In this section, the max-p-regions method and the random-regions method are evaluated on Germany, Spain, France and Italy. The evaluation procedure for each country begins by running both spatial resolution methods using the same minimum threshold area value. In most cases, both generated the same number of regions. If their sum of regions generated was less than ten but greater than two, a power system optimisation was conducted. The upper limit of the number regions per power

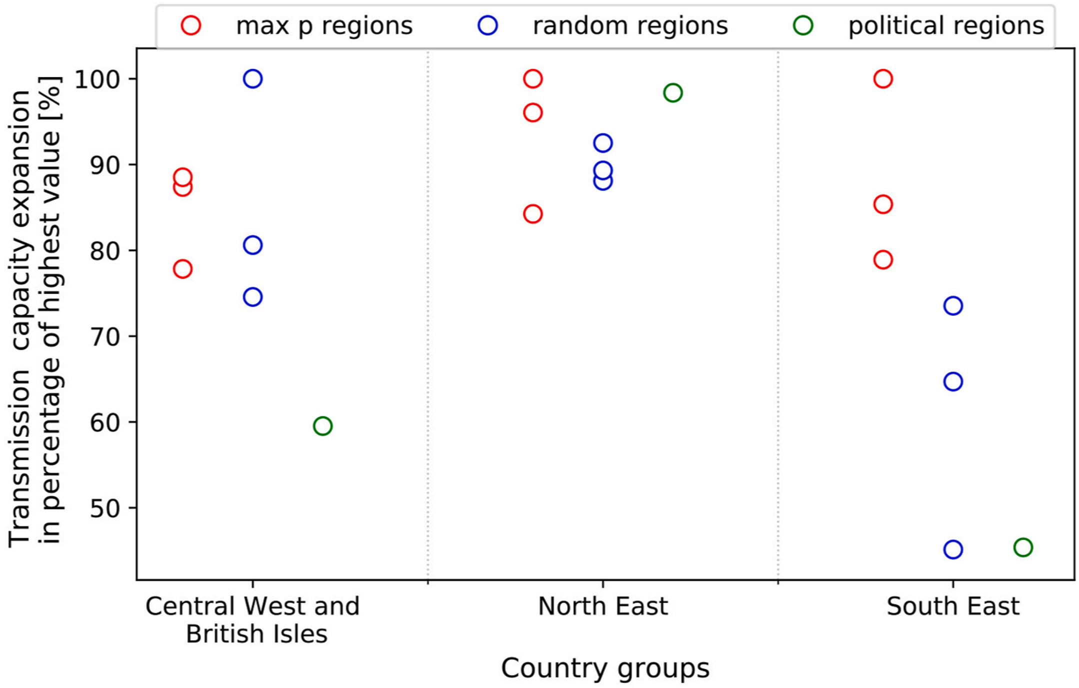


FIGURE A.4: The sums of the transmission line capacity expansions obtained from the optimisation results of the regions sets, given in percentage of the highest value in the same country group.

system was set to minimise the computational hours of the exercise. Whereas, the lower limit is the expected minimum number of regions types classified with the max-p-regions method.

The minimum threshold value was altered to create power system models with a different number of regions, for both spatial resolution reduction methods. As discussed in Section 2, the initial seed value used can impact how the regions are defined. Thus, the spatial resolution was conducted three times with different seed values, for each spatial resolution reduction method, to give a reasonable statistical representation of the variation in resulting regions. The sum of the solar and wind capacity expansion and transmission capacity expansion values, from the optimisation results, have been plotted in Fig. A.5. The same approach used in Fig. A.3, Fig. A.4 is used in Fig. A.5 to evaluate how the effects of spatial resolution reduction impact the optimisation results of the individual sets of regions.

The result for Germany and Italy (Fig. A.5) shows that on average, the max-p-regions generated lower solar and wind capacity expansion values and higher transmission expansion values than the random regions. In the case of Germany, the highest three number of regions, 7 to 9, show similar results for both spatial resolution reduction effects. There is a more pronounced difference between the results of the number of regions 6 and 7. Values for Italy display the same trends, where

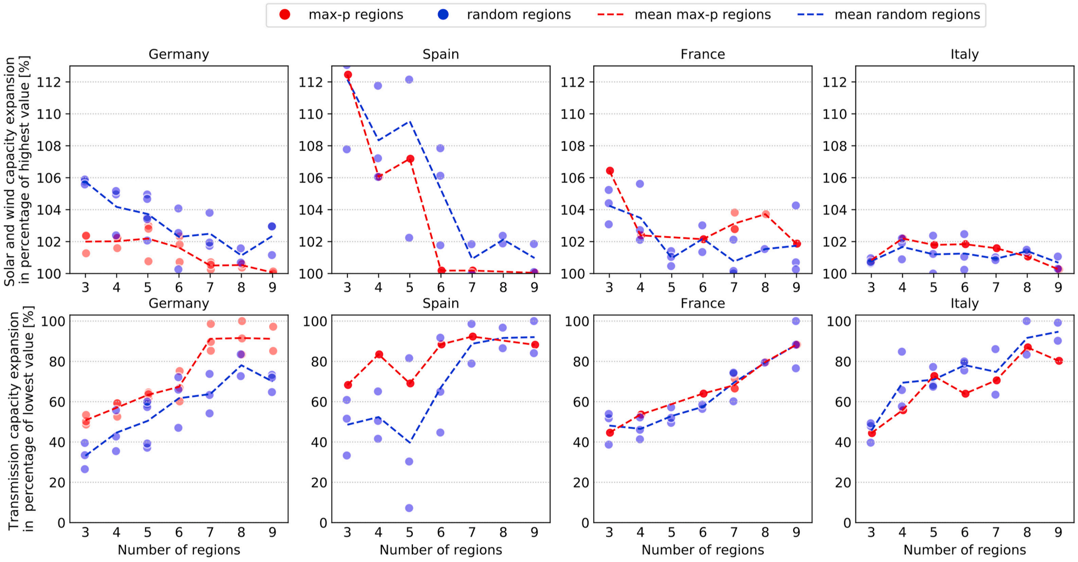


FIGURE A.5: The sums of the solar and wind capacity expansion and transmission capacity expansion values are given in percentage of the highest and lowest value respectively. These percentage values are from the optimisation results of power systems of Germany, Spain, France and Italy, after multiple spatial resolution reduction, using max-p-regions and random-regions method.

the highest four number of regions have similar results. However, there is a clear difference between the values of the number of regions 5 and 6. This observation could indicate that during spatial resolution reduction, certain incremental changes to the magnitude of spatial resolution selected could have greater impacts on the results than others. The average values from Italy and France do not indicate any clear difference between the optimisation results of the max-p-regions and the random regions. In general, there are two clear trends shown in Fig. A.5. First, the solar and wind expansion capacity values increase with the decrease in the number of regions. Secondly, the transmission capacity expansion values decrease with the decrease in the number of regions. Both trends support the assumptions made about the effects of spatial resolution reduction on ESOMs.

A.5 Discussions and conclusions

This paper presents a targeted approach to define regions during spatial resolution reduction called the max-p-regions method. The max-p-regions method clustered areas into regions based on their similarity in terms of their electricity consumption, solar and wind generation potential and energy storage capabilities. The regions generated by the max-p-regions method were compared with two other spatial resolution reduction methods. A comparative evaluation was conducted to compare the impacts of the spatial resolution reduction methods on energy system optimisation models.

Results from the comparative evaluation on country groups showed that the use of the max-p-regions method to define regions resulted in the optimisation results of the ESOMs to be less affected by the effects of spatial resolution reduction in comparison to using national jurisdiction to define regions. Particularly for the group of countries located in the South Eastern region of Europe, the set of regions generated by the max-p-regions method had on average less solar and wind capacity expansion and higher transmission capacity expansion than the other two spatial resolution reduction methods. In the case of Germany and Spain, the max-p-regions method was more effective in identifying sets of regions, which were less impacted by the two recognised effects of spatial resolution reduction, in comparison to using the non-targeted random approach. In contrast, the results did not indicate any benefits to using the max-p-regions method in comparison to the non-targeted random approach when applied to France and Italy.

Ideally, the optimisation results of the power system models before applying the spatial resolution reduction methods would be used as the reference to investigate the effectiveness of the max-p-regions at minimising the spatial resolution effects in comparison to the other methods. Unfortunately, computational limitations did not allow for this ideal method of evaluation. The alternative approach used in this study was to use observations made by previous studies as indicators to measure the effects of spatial resolution reduction. The observations made from varying the spatial resolution of the ESOMs of the individual countries supported the alternative approach undertaken. These observations showed that the solar and wind capacity expansion increases and the transmission capacity expansion decreases as the spatial resolution of ESOMs are reduced.

In conclusion, the effectiveness of using the max-p-regions method to minimise the effects of spatial resolution reduction on ESOMs appears to be specific to the spatial scope. Future studies could investigate what characteristics of the spatial scope impact the effectiveness of the max-p-regions method. It would be interesting to examine whether the max-p-regions method could be improved or adjusted to take into account the specificity of the spatial scope being analysed. One possible improvement would be to use a different attribute to determine the minimum threshold value of the max-p-regions method.

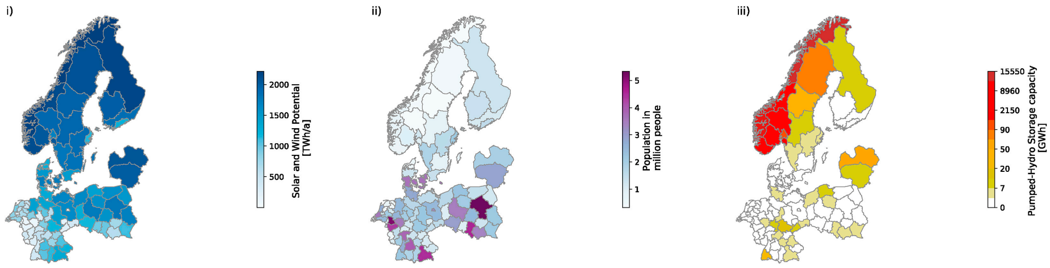
Declaration of competing interest

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

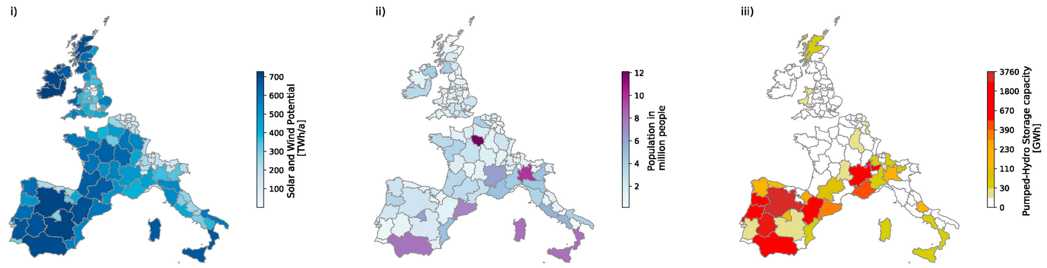
Acknowledgements

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.

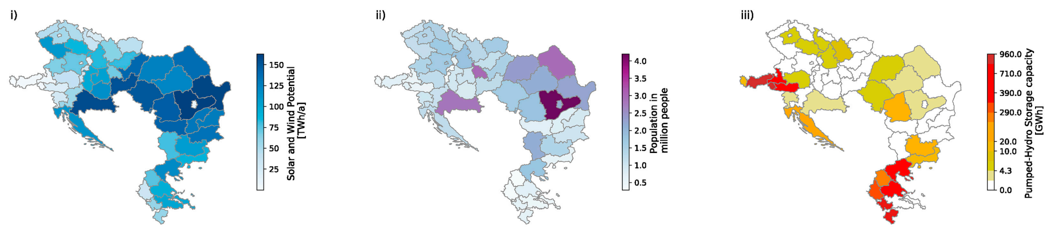
A.A Distribution of the spatial attributes used in the max-p-regions method across the Nomenclature of territorial units for statistics level 2 (NUTS 2) for country groups, North East, Central West and British Isles, and South East



(a) The spatial distribution of i) solar and wind potential, ii) population and iii) pumped-hydro storage capacity across the NUTS 2 areas of North East.

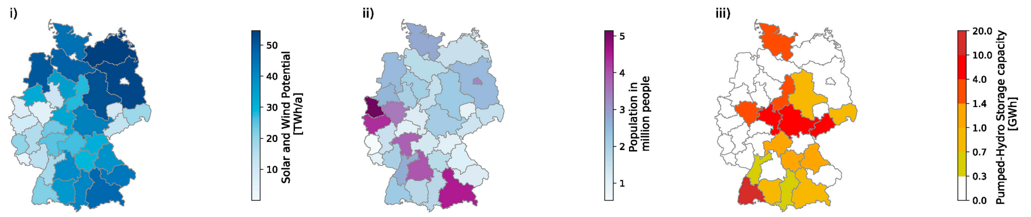


(b) The spatial distribution of i) solar and wind potential, ii) population and iii) pumped-hydro storage capacity across the NUTS 2 areas of Central West and British Isles.

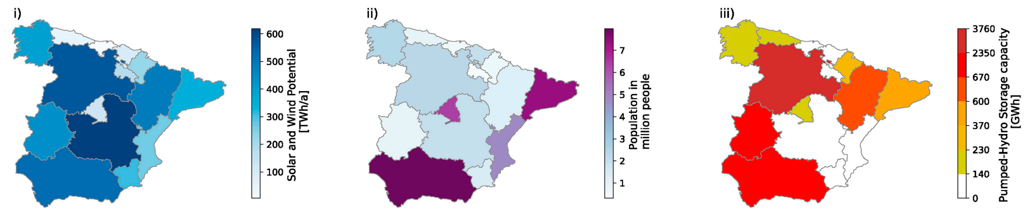


(c) The spatial distribution of i) solar and wind potential, ii) population and iii) pumped-hydro storage capacity across the NUTS 2 areas of South East.

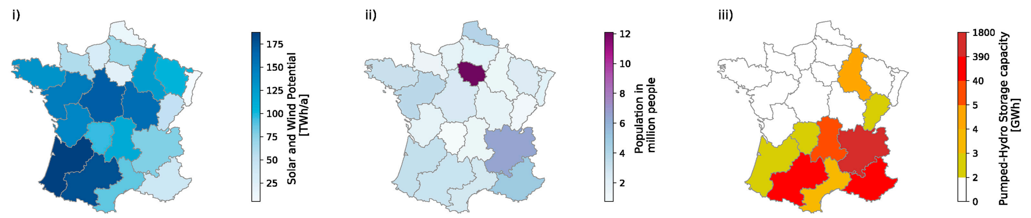
A.B Distribution of the spatial attributes used in the max-p-regions method across the Nomenclature of territorial units for statistics level 2 (NUTS 2) for, Germany, Spain, France and Italy. For Italy the island NUTS 2 areas of Sardinia and Sicily is grouped with the Calabria area, such that the values represented for each area represents the values of the grouped area



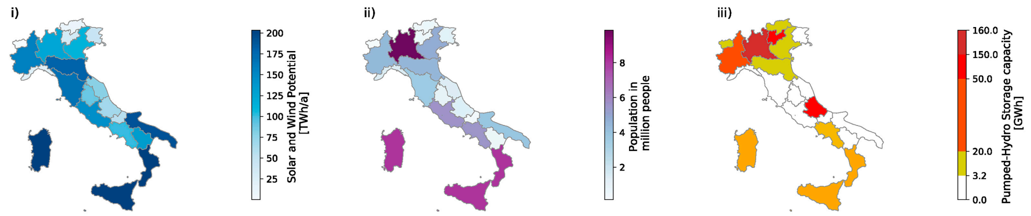
(a) The spatial distribution of i) solar and wind potential, ii) population and iii) pumped-hydro storage capacity across the NUTS 2 areas of Germany.



(b) The spatial distribution of i) solar and wind potential, ii) population and iii) pumped-hydro storage capacity across the NUTS 2 areas of Spain.

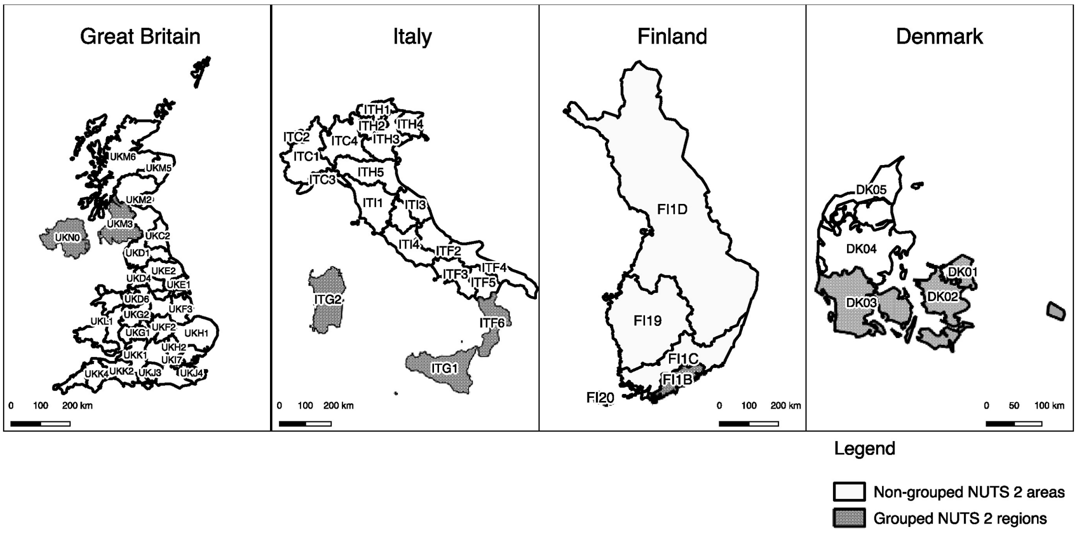


(c) The spatial distribution of i) solar and wind potential, ii) population and iii) pumped-hydro storage capacity across the NUTS 2 areas of France.

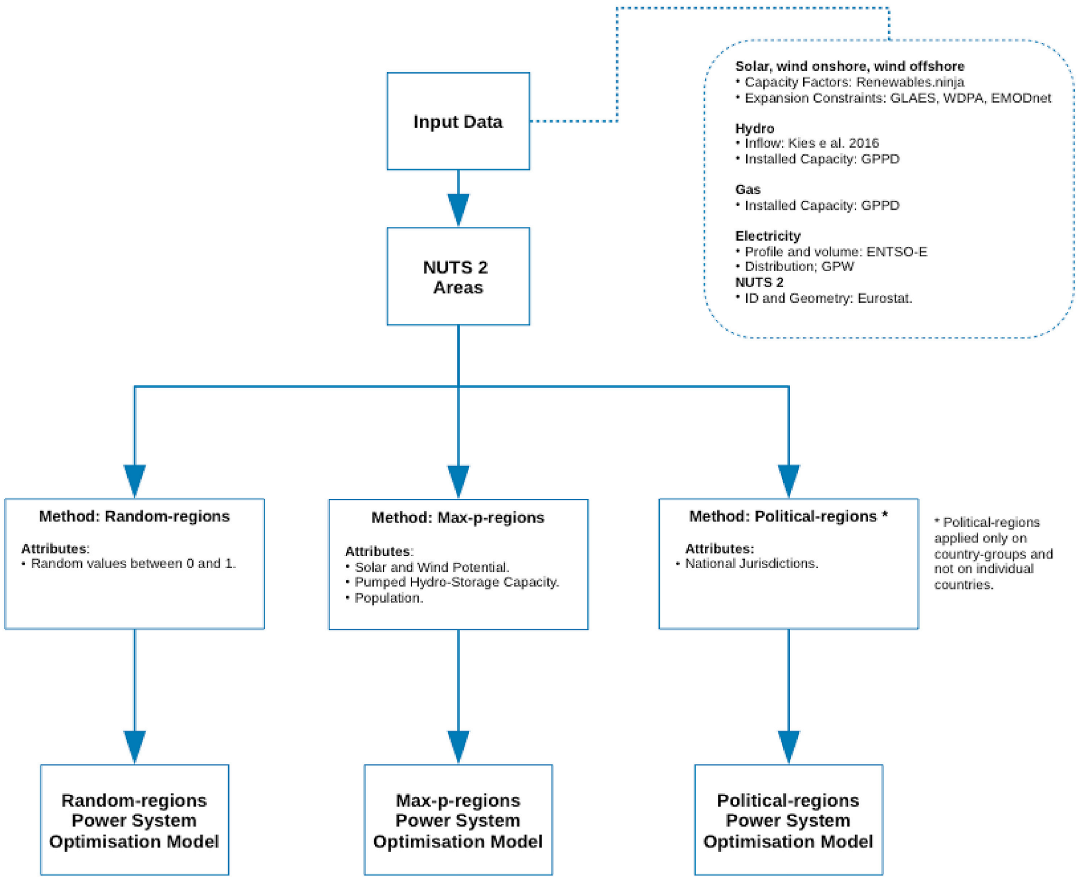


(d) The spatial distribution of i) solar and wind potential, ii) population and iii) pumped-hydro storage capacity across the NUTS 2 areas of Italy.

A.C Graphical illustration of countries that have island NUTS 2 areas that are grouped with a continental NUTS 2 area



A.D Flow diagram showing the connections between the different processes used to build the power system optimisation models



A.E DC Links between NUTS 2 areas

Id	Name	from	To	Installed Capacity	Planned capacity
				MW	MW
1	Estlink	FI1B	EE00	350	350
2	Estlink 2	FI1B	EE00	650	650
3	Kriegers Flak	DE80	DK02	0	400
4	Baltic	DE80	SE22	600	600
5	Kontek	DE80	DK02	600	600
6	Konti-Skan (12)	DK02	SE23	550	550
7	COBRA cable	DK03	NL11	0	700
8	Skagerrak	DL05	NO04	1700	1700
9	Nordlink	DEF0	NO04	0	1400
10	Swedlink	SE21	LT00	700	700
11	NorNed	NO04	NL11	700	700
12	Nord Sea Link	NO04	UKC2	0	1400
13	Nord Connect	NO04	UKM5	0	1400
14	Viking Link	DK03	UKF3	0	1400
15	BritNed	UKJ4	NL32	1000	1000
16	Nemo Link	BE25	UKJ4	1000	1000
17	GridLink	UKJ4	FR30	1400	1400
18	ElecLink	UKJ4	FR30	0	1000
19	Aquind	UKJ4	FR30	0	2000
20	IFA 2	UKJ3	FR25	0	1000
21	IFA Link	UKK4	FR25	0	1400
22	IFA 1	UKJ4	FR30	2000	2000
23	Celtic Link	FR52	IE02	0	700
24	Green Link	IE02	UKL1	0	500
25	East-West Link	IE02	UKD6	500	500
26	Moyle	UKN0	UKM3	500	500
27	Western Link	UKM3	UKD6	2200	2200
28	Italy-Greece	ITF4	EL54	500	500
29	SACOI	ITG2	ITI1	300	300
30	SAPEI	ITG2	ITI4	1000	1000

References

[1] Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kobayashi, S., Kriegler, E., Mundaca, L., Séférian, R., Vilariño, M. V., Calvin, K., Emmerling, J., Fuss, S., Gillett, N., He, C., Hertwich, E., Höglund-Isaksson, L., Huppmann, D., Luderer, G., McCollum, D. L., Meinshausen, M., Millar, R., Popp, A., Purohit, P., Riahi, K., Ribes, A., Saunders, H., Schädel, C., Smith, P., Trutnevyte, E., Xiu, Y., Zhou, W., Zickfeld, K., Flato, G., Fuglestvedt, J., Mrabet, R., and Schaeffer, R. "Mitigation Pathways Compatible with 1.5 °C in the Context of Sustainable Development". en. In: (2018), p. 82.

[2] Hörsch, J. and Brown, T. "The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios". In: *2017 14th International Conference on the European Energy Market (EEM)*. ISSN: 2165-4093. June 2017, pp. 1–7.

[3] Pfenninger, S., Hawkes, A., and Keirstead, J. "Energy systems modeling for twenty-first century energy challenges". en. In: 33 (May 2014), pp. 74–86.

[4] Wohland, J., Brayshaw, D. J., and Pfenninger, S. "Mitigating a century of European renewable variability with transmission and informed siting". In: (May 2021).

[5] Heaps, C. "USER GUIDE FOR LEAP 2008". en. In: (2010), p. 318.

[6] Howells, M., Rogner, H., Strachan, N., Heaps, C., Huntington, H., Kypreos, S., Hughes, A., Silveira, S., DeCarolis, J., Bazillian, M., and Roehrl, A. "OSeMOSYS: The Open Source Energy Modeling System". en. In: *Energy Policy* 39.10 (Oct. 2011), pp. 5850–5870.

[7] Østergaard, P. A. "Reviewing EnergyPLAN simulations and performance indicator applications in EnergyPLAN simulations". en. In: *Applied Energy* 154 (Sept. 2015), pp. 921–933.

[8] Frew, B. A. and Jacobson, M. Z. "Temporal and spatial tradeoffs in power system modeling with assumptions about storage: An application of the POWER model". en. In: 117 (Dec. 2016), pp. 198–213.

[9] Krishnan, V. and Cole, W. "Evaluating the value of high spatial resolution in national capacity expansion models using ReEDS". In: *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, July 2016, pp. 1–5.

[10] Simoes, S., Zeyringer, M., Mayr, D., Huld, T., Nijs, W., and Schmidt, J. "Impact of different levels of geographical disaggregation of wind and PV electricity generation in large energy system models: A case study for Austria". en. In: 105 (May 2017), pp. 183–198.

[11] Anderski, T., Surmann, Y., Stemmer, S., Grisey, N., Momot, E., Leger, A.-C., Betraoui, B., and Van Roy, P. *D2.2 European cluster model of the Pan European transmission grid*. 2015.

[12] Siala, K. and Mahfouz, M. Y. "Impact of the choice of regions on energy system models". en. In: 25 (Aug. 2019), pp. 75–85.

[13] Duque, J. C., Anselin, L., and Rey, S. J. "THE MAX-P-REGIONS PROBLEM*". en. In: *Journal of Regional Science* 52.3 (Aug. 2012), pp. 397–419.

[14] Rey, S. J. and Anselin, L. "PySAL: A Python Library of Spatial Analytical Methods". en. In: *Handbook of Applied Spatial Analysis*. Ed. by M. M. Fischer and A. Getis. Springer Berlin Heidelberg, 2010, pp. 175–193.

[15] Walt, S. v. d., Colbert, S. C., and Varoquaux, G. "The NumPy Array: A Structure for Efficient Numerical Computation". In: *Computing in Science Engineering* 13.2 (Mar. 2011), pp. 22–30.

[16] Brown, T., Hörsch, J., and Schlachtberger, D. "PyPSA: Python for Power System Analysis". en. In: *Journal of Open Research Software* 6 (Jan. 2018). arXiv: 1707.09913.

[17] European Commission, Eurostat (ESTAT), GISCO. *Nomenclature of Territorial Units for Statistics (NUTS) 2013 - Statistical Units - Data set*. Type: dataset. Dec. 2015.

[18] ENTSO-E. *Detailed Data Descriptions*. European Network of Transmission System Operators for Electricity. 2014.

[19] *Gridded Population of the World (GPW), v4 | SEDAC*.

[20] Byers, L., Friedrich, J., Hennig, R., Kressig, A., Li, X., Mccormick, C., and Valeri, L. M. "A GLOBAL DATABASE OF POWER PLANTS". en. In: (), p. 18.

[21] Pfenninger, S. and Staffell, I. "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data". en. In: *Energy* 114 (Nov. 2016), pp. 1251–1265.

[22] Kies, A., Kabitri Chattopadhyay, Elke Lorenz, Lüder von Bremen, and Detlev Heinemann. *Simulation of renewable feed-in for power system*. Tech. rep. Institute of Physics, University of Oldenburg, May 2016.

[23] Ryberg, D. S., Robinius, M., and Stolten, D. "Methodological Framework for Determining the Land Eligibility of Renewable Energy Sources". en. In: (), p. 35.

[24] *Pipelines | EMODnet Human Activities*.

[25] E. E. Agency, ed. *Greenhouse gas emissions in Europe: a retrospective trend analysis for the period 1990 - 2008*. en. EEA report 2011,6. OCLC: 799048672. Luxembourg: Publ. Off. [u.a.], 2011.

[26] Scarlat, N., Dallemand, J.-F., and Fahl, F. "Biogas: Developments and perspectives in Europe". en. In: *Renewable Energy* 129 (Dec. 2018), pp. 457–472.

[27] *JRC Hydro-power plants database. Contribute to energy-modelling-toolkit/hydro-power-database development by creating an account on GitHub*. original-date: 2019-04-05T13:38:23Z. Apr. 2019.

[28] Jülch, V. "Comparison of electricity storage options using levelized cost of storage (LCOS) method". en. In: *Applied Energy* 183 (Dec. 2016), pp. 1594–1606.

[29] Energienet and Danish Energy Agency. *Technology Data for Energy Plants for Electricity and District heating generation*. Tech. rep. 0009. Aug. 2016.

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

Paper B
Christian Etienne Fleischer [a]
[a] Department of Energy and Environmental Management, Europa-Universität Flensburg, 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.

Keywords: Max-p-regions, Spatial aggregation, Renewables

Method name: Defining regions according to energy-related attributes using the max-p-regions method for energy system modelling

Paper published in MethodX Volume 8 (2021)
https://doi.org/10.1016/j.mex.2021.101211

Specifications table

Subject area:	Energy
More specific subject area:	Energy system modelling
Method name:	Defining regions according to energy-related attributes using the max-p-regions method for energy system modelling.
Name and reference of original method:	The regionalisation method is a combination of the European cluster model method [1] and the spatial aggregation clustering method detailed in [2].
Resource availability:	Data • Nomenclature of territorial units for statistics level geometry [3] • Gridded population of the world version 4 [4] • JRC hydro-power plants database [5] • Normalized maximum power output profiles [21] • Offshore oil and gas pipelines [6] • Protected areas [7] Software • Geospatial Land Availability for Energy Systems framework (GLAES) [8] • pysal [9] • numpy [10] • geopandas [11] • rasterstats [12]

Method details

B.1 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 [13], [14], [15]. The max-p regions method presented in [16] 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 approached 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 [17]. The random regions method described in this article is used in [16] to verify the efficacity of the max-p regions method to define regions that are less impacted by spatial resolution reduction effects.

B.2 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 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.

The sum of the electricity storage capacity of pumped storage hydropower plants within the NUTS 2 areas are added to the spatial dataset. The information on pumped storage hydropower plants is from the Joint Research Centre (JRC) hydropower plants database [5]. The JRC hydropower plants database pumped storage hydropower plants are labelled HPHS in the database type column. The longitude and latitude location values of the hydropower plant database are in CRS EPSG 4326 and therefore are first converted to CRS EPSG 3035. The sum of the electricity storage capacity of pumped storage hydropower plants within the NUTS 2 areas is the sum of the storage capacity of the HPHS labelled hydropower plants located within the NUTS 2 areas. For HPHS labelled hydropower plants whose storage capacity is not given or given as zero, the storage capacity is assumed to be six times the installed power capacity of the plant. The last set of data attributed to NUTS 2 area dataset is the electricity generation potential from wind and solar technology available in the NUTS 2 areas. Two factors determine the electricity generation potential. The first factor is the installation capacity potential assigned to that region which is dependent on the eligible land or offshore area assigned to the NUTS 2 area.

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/km^2, 4 MW/km^2 and 12 MW/km^2 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 [18]. 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 areas. The technology with the highest electricity generation potential is assigned to the NUTS 2 area dataset.

B.3 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 B.1 has details of the islands and the continental NUTS 2 areas.

TABLE B.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

B.4 Defining regions using the max-p-regions method

The Python Spatial Analysis Library (PySAL) contains a spatial optimisation library called spopt [9]. As described in the pseudocode 1 below, the MaxPHeuristic function of the spopt library, in conjunction with the libpysal python library, is used to implement the max-p-regions method.

The MaxPHeuristic 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 [17], there are three options to conduct a local search when finding the best feasible solution to define regions. The MaxPHeuristic function uses the simulate annealing approach. A contiguity constraint is defined in the MaxPHeuristic function by assigning weight values to the NUTS 2 areas using the queen contiguity weights function provided by libpysal [19]. The max-p regions method is detailed in Fig. B.1.

The result of the MaxPHeuristic 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 spatial attributes are assigned as the features attribute in the MaxPHeuristic function to create energy-related max-p regions.

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: Max-p regions method

gdp: Geopandas dataframe of NUTS 2 areas without islands,
A: Set of areas in gdp,
W: Neighborhoods ,
l: 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.

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

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

pseudocode 1 in Fig. B.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 [16] 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. B.2.

The complete process to create the energy-related max-p regions and the random regions is illustrated in Fig. B.3.

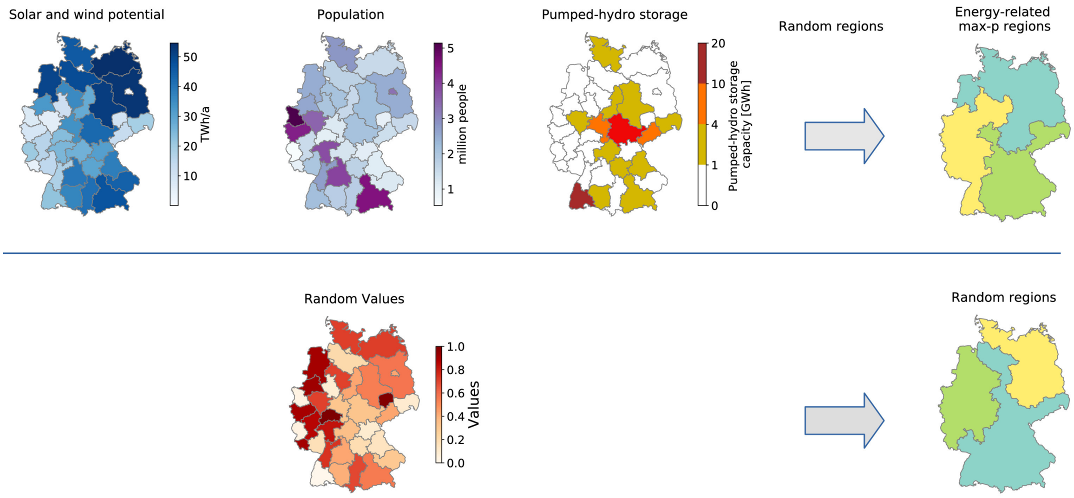


FIGURE B.2: Example of creating three energy-related max-p regions and three random regions from the NUTS 2 areas of Germany. The energy-related attributes solar and wind potential, population and pumped-hydro storage are used to determine the heterogeneity between regions and create energy-related max-p regions. Random values from 0 to 1 and are used to determine the heterogeneity between regions to create random regions.

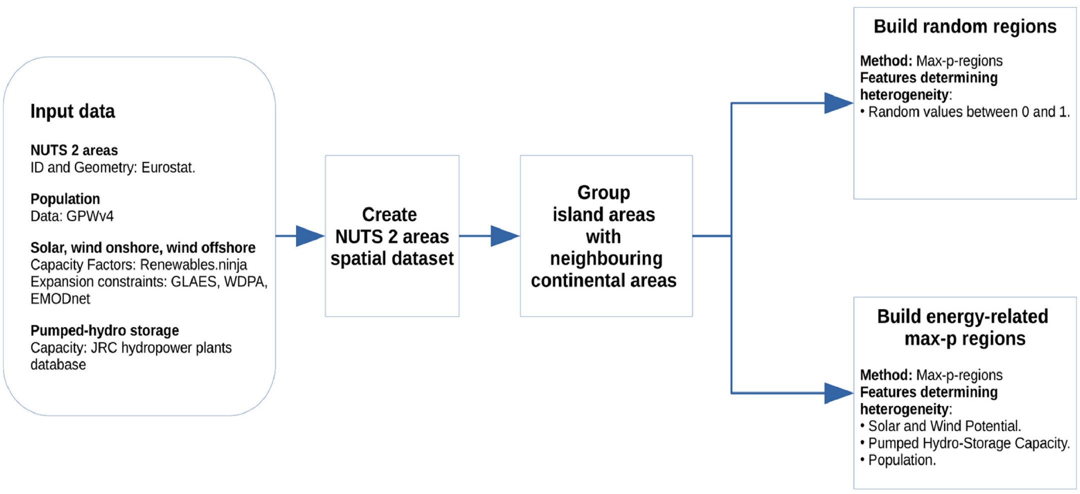


FIGURE B.3: Illustration of the process to create energy-related max-p regions and random regions.

B.5 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] Anderski, T., Surmann, Y., Stemmer, S., Grisey, N., Momot, E., Leger, A.-C., Betraoui, B., and Van Roy, P. *D2.2 European cluster model of the Pan European transmission grid*. 2015.

[2] Siala, K. and Mahfouz, M. Y. "Impact of the choice of regions on energy system models". en. In: 25 (Aug. 2019), pp. 75–85.

[3] Geographic Information System of the COmmission. *GISCO data distribution API - NUTS 2016*.

[4] *Gridded Population of the World (GPW), v4 | SEDAC*.

[5] *JRC Hydro-power plants database. Contribute to energy-modelling-toolkit/hydro-power-database development by creating an account on GitHub*. original-date: 2019-04-05T13:38:23Z. Apr. 2019.

[6] *Pipelines | EMODnet Human Activities*.

[7] *Protected Planet*.

[8] Ryberg, D. S., Robinius, M., and Stolten, D. "Methodological Framework for Determining the Land Eligibility of Renewable Energy Sources". en. In: (), p. 35.

[9] Rey, S. J. and Anselin, L. "PySAL: A Python Library of Spatial Analytical Methods". en. In: *Handbook of Applied Spatial Analysis*. Ed. by M. M. Fischer and A. Getis. Springer Berlin Heidelberg, 2010, pp. 175–193.

[10] Walt, S. v. d., Colbert, S. C., and Varoquaux, G. "The NumPy Array: A Structure for Efficient Numerical Computation". In: *Computing in Science Engineering* 13.2 (Mar. 2011), pp. 22–30.

[11] Jordahl, K., Bossche, J. V. den, Wasserman, J., McBride, J., Fleischmann, M., Gerard, J., Tratner, J., Perry, M., Farmer, C., Hjelle, G. A., Gillies, S., Cochran, M., Bartos, M., Culbertson, L., Eubank, N., Bilogur, A., and maxalbert. *geopandas/geopandas: v0.7.0*. Feb. 2020.

[12] *rasterstats — rasterstats 0.10.3 documentation*.

[13] Krishnan, V. and Cole, W. "Evaluating the value of high spatial resolution in national capacity expansion models using ReEDS". In: *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, July 2016, pp. 1–5.

[14] Hörsch, J. and Brown, T. "The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios". In: *2017 14th International Conference on the European Energy Market (EEM)*. ISSN: 2165-4093. June 2017, pp. 1–7.

[15] Frew, B. A. and Jacobson, M. Z. "Temporal and spatial tradeoffs in power system modeling with assumptions about storage: An application of the POWER model". en. In: 117 (Dec. 2016), pp. 198–213.

[16] Fleischer, C. E. "Minimising the effects of spatial scale reduction on power system models". en. In: *Energy Strategy Reviews* 32 (Nov. 2020), p. 100563.

[17] Duque, J. C., Anselin, L., and Rey, S. J. "THE MAX-P-REGIONS PROBLEM*". en. In: *Journal of Regional Science* 52.3 (Aug. 2012), pp. 397–419.

[18] Pfenninger, S. and Staffell, I. "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data". en. In: *Energy* 114 (Nov. 2016), pp. 1251–1265.

[19] Brown, T., Hörsch, J., and Schlachtberger, D. "PyPSA: Python for Power System Analysis". en. In: *Journal of Open Research Software* 6 (Jan. 2018). arXiv: 1707.09913.

A data processing approach with built-in spatial resolution reduction methods to construct energy system models

Paper C

Christian Etienne Fleischer [a]
[a] Department of Energy and Environmental Management, Europa-Universität Flensburg, Flensburg, Germany

Abstract

Introduction: Data processing is a crucial step in energy system modelling which prepares input data from various sources into a format needed to formulate a model. Multiple open-source web-hosted databases offer pre-processed input data within the European context. However, the number of documented open-source data processing workflows that allow for the construction of energy system models with specified spatial resolution reduction methods is still limited.

Methods: The first step of the data-processing method builds a dataset using web-hosted pre-processed data and open-source software. The second step aggregates the dataset using a specified spatial aggregation method. The spatially aggregated dataset is used as input data to construct sector-coupled energy system models.

Results: To demonstrate the application of the data processing process, three power and heat optimisation models of Germany were constructed using the proposed data processing approach. Significant variation in generation, transmission and storage capacity of electricity were observed between the optimisation results of the energy system models.

Conclusions: This paper presents a novel data processing approach to construct sector-coupled energy system models with integrated spatial aggregations methods.

Paper published in Open Research Europe (2021)
https://doi.org/10.12688/openreseurope.13420.1

C.1 Introduction

In the past, energy system models were primarily closed and proprietary. However, recently more open-source energy system modelling tools have been made available. Maruf et al. identified 59 freely available energy system modelling tools [1]. Energy system models are considered “open” when the data and model code is accessible and legally usable [2]. Pfenninger et al. discuss how open models improve the scientific quality of the models by adhering to fundamental scientific principles such as transparency and reproducibility [3]. Pfenninger et al. also state that when models and data are open, productivity increases as it reduces the time spent by researchers in duplication of work in developing models and datasets [3].

The steps in the open-source energy modelling process are described by Pfenninger et al. in [2]. One crucial step in that process is data processing. Data processing is an intermediate step between the raw input data and the model formulation. The input data is made accessible to the formulated model after undergoing data processing. The methods used to process the input data can have an impact on model results. Two documented impacts are the effects of temporal resolution reduction methods [4–9] and spatial resolution reductions methods[10–14]. Therefore, the data processing steps must be well documented to ensure that their impact on the modelling results can be properly gauged. There are a limited amount of available open-source modelling tools and datasets that allow for the alteration of the spatial resolution of energy system models. One of these tools is presented in [15] by Hörsch et al., which builds a highly spatially disaggregated European power system model dataset. The resolution of the dataset can then be reduced at various spatial scale by clustering the electrical network using the k-means algorithm. Tröndle et al. investigate the possibility of renewable energy autarky models at four different spatial scales: continental level, national level, regional level and municipal level [16].

Input data of high spatial and temporal resolution can be generated using tools such as the global Renewable Energy atlas (REatlas) atlas [17], the Python Generator of renewable time series and maps (PyGreta) [18], and the GlobalEnergyGIS [19]. The input data can also be obtained from an extensive list of web-hosted platforms, repositories and databases. These platforms and datasets include the renewables.ninja platform [20], the Open Power System Data (OPSD) platform [21], the hotmaps repository [22] and the ENSPRESO database [23]. The Open Energy Platform compliments these platforms by documenting and sharing datasets used by existing energy system models such as the eTraGo [24], OSeMBE [25] and MEDEAS [26]. These platforms facilitate the identification of documented and validated input data in centralised locations.

This paper presents a novel data processing approach that maximises the use of the web-hosted pre-processed input data to build energy system models. The appliction of the data processing approach is demonstrated by building three power and heat models with different spatial contexts. The differences between the spatial contexts are the spatial scope and spatial zones that define the regions in the models.

C.2 Methods

C.2.1 Data processing workflow approach

The proposed data processing approach can be split into two steps as illustrated in Figure C.1. The first step builds the Areas dataset, which host the necessary data variables from the pre-processed input data sources within a structured framework. A set of requirements are defined for the Areas dataset to ensure standardisation and proper documentation of the data variables. The first requirement prescribes that the data variables need to be indexed using standardised reference keys. These reference keys allow the data to be uniformly organised according to spatial, temporal, and technological specifications. The Nomenclature of Territorial Units for Statistics level 2 (NUTS 2) and NUTS level 0 are the two spatial reference keys used by the Areas dataset to structure data with a spatial dimension. The second requirement ensures the use of standardised units of measurements. The standardisation of units ensures uniformity, allowing the use of the dataset for modelling without additional unit conversions. The final requirement for building an Areas dataset is the documentation of the data variables in the dataset. The documentation entails providing the source of the data variables and a description of the unit.

Building a Regions dataset is the second step of the data processing process. The Regions dataset is constructed by first defining the areas of interest. The areas of interest could include all areas in the Areas dataset, or it could consist of a subset of areas. As the name of the dataset indicates, the Regions dataset allows for the grouping of areas into regions. The Regions dataset is described in more detail in Regions dataset.

Dataset framework. Two requirements were defined to guide the selection process of the dataset framework used as the skeleton for the Areas and the Regions dataset. The first requirement is that the dataset framework should be able to handle data with more than one reference key. The need for multiple reference keys handling capability allows data variables to be referenced according to spatial, temporal, or

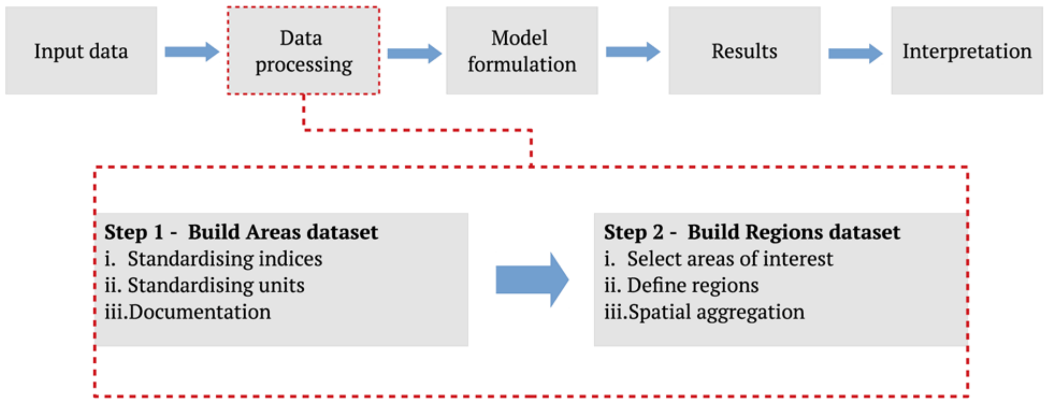


FIGURE C.1: An illustrated description of the proposed data processing workflow of the energy system modelling process.The energy system modelling process is based on the Openmod Philosophy laid out by Pfenninger et al. in [2].

even technological indices. The second requirement is that the dataset framework must integrate well with existing open-source energy modelling software and scientific analysis software. As there is a multitude of these software [15, 27–30] written in the Python programming language [31], the dataset framework should be a Python-based software. The xarray dataset object in the xarray toolkit [32] was selected to build the datasets based on these two requirements.

There are additional benefits of using xarray to construct the datasets, as listed below:

- the dataset can be exported as a unidata network common (nc) file format that can be compressed to lower file sizes which eases sharing of the datasets;
- the dataset framework allows for documentation of the data variables.

The datasets created can be stored and shared using a single file in online archives such as Zenodo. Zenodo attaches Digital Object Identifiers (DOIs), which allows for the citation of the data. Hörsch et al.[15] and Tröndle et al.[16] both share different versions of their model on Zenodo using the nc file format.

Areas dataset. The Areas dataset spatial scope includes the EU 27 countries with the exclusion of Cyprus and Malta and the addition of Norway, Great Britain, and Switzerland. The data variables in the dataset can be sub-divided into two sets, the base variables and the derived variables. The base variables are used to determine the derived variables.

The Areas and the Regions datasets have a total of seven reference keys. There are two spatial reference keys, namely NUTS 0 and NUTS 2. Data variables using the NUTS 0 reference key apply at the national level, whereas the NUTS 2 level data are, for most countries, spatial administrative areas within the countries. There are four technology-specific reference keys: techs, fuels, techs hydro, and techs hydro subset. The techs and fuels reference keys reference the different technology types of the power plants variable. The techs hydro reference key is used to reference the technologies used to generate power from hydropower. The techs hydro subset dimension refers to a subset of the techs hydro reference key of hydropower technologies that have reservoirs to store water. The time reference key is used to index the time dimension of data variables. The time indices are the hourly intervals of a single representative year. The variables reference key, unit and sources are summarised in Table C.1. Except for temperature and offshore wind capacity factor, the data variables in the Areas dataset are organised at the NUTS 2 spatial level. Therefore, the Areas dataset can be considered as a collection of data variables of 270 NUTS 2 areas.

Base variables

There are a set of base variables needed to establish the foundation of the dataset. The identification code (NUTS 2 id), the geometrical information (Geometry), and the country identification code (Country code) of the NUTS 2 areas are base variables taken from the Eurostat database [33]. The NUTS 2 geometry scale is 1:10 million. The ambient air temperature obtained from annual hourly historical weather data from the renewables.ninja platform [20] is a base variable in the NUS T 2 area dataset. The weather data is only available at the country level, so the temperature data is given at NUTS 0 in the dataset. The population values in the NUTS 2 areas are extracted from the Global Human Settlement (GHS) population grid [35]. The population raster for the year 2015 was used. The population values within the geometry of the NUTS 2 area are summed, and the sum is the population value assigned to the NUTS 2 area. Other base variables describe the energy generation and storage technologies.

The power plants data variable gives the aggregated installed capacity of conventional power plants, solar photovoltaic (PV) installations, onshore and offshore wind installations associated with the NUTS 2 areas. Information on conventional power plants is from the conventional power plants dataset hosted on the OPSD platform [38]. Information on solar and onshore wind installations for NUTS 2 areas in Germany, Denmark, France, Poland, United Kingdom and Switzerland, were extracted from the OPSD renewable power plants dataset [36]. A power plant from the two

TABLE C.1: Summary of the data variables in the Areas dataset.

Data variable	Base/ Derived variables	Reference key	Unit	Sources
NUTS 2 id	Base	*NUTS 2*	NA	[33]
Geometry	Base	*NUTS 2*	NA	[33]
Country code	Base	*NUTS 2*	NA	[33]
Temperature	Base	*time, NUTS 2*	Degrees Celsius	[34]
Population	Base	*NUTS 2*	people	[35]
Power plants	Base	*techs, fuels, NUTS 2*	MW	[36–38]
Hydropower plants dispatch	Base	*techs hydro, NUTS 2*	MW	[39]
Hydropower plants storage	Base	*techs hydro subset, NUTS 2*	MW, MWh	[39]
Rooftop solar PV area	Base	*NUTS 2*	km^2	[40–43]
Ground-mounted solar PV	Base	*NUTS 2*	km^2	[40–43]
Onshore wind area	Base	*NUTS 2*	km^2	[22, 44, 45]
Offshore wind area	Base	*NUTS 2*	km^2	[22, 44–46]
Solar capacity factor	Base	*time, NUTS 2*	Per unit capacity	[20]
Wind capacity factor	Base	*time, NUTS 2*	Per unit capacity	[20]
Offshore wind capacity factor	Base	*time, NUTS 0*	Per unit capacity	[20]
Hydropower capacity factor	Base	*time, NUTS 2*	Per unit capacity	[47]
Power	Derived	*time, NUTS 0*	MW	[48]
Heat	Derived	*time, NUTS 2*	MW	[22]
Air-source apacity factor	Derived	*time, NUTS 2*	Per unit capacity	[20, 22]

datasets is associated with a NUTS 2 area when the power plants' geometrical information places it within the geometrical boundaries of that NUTS 2 area. The power plants are grouped by the fuel and technology type and aggregated by their installed generating capacity in megawatts (MW). As the offshore wind installations are not within the geometries of the NUTS 2 areas, they are determined separately. The existing offshore wind installations were extracted from the European Marine Observation and Data Network (EMODnet) offshore wind farm database [37]. This database provides the location of each wind farm as a georeferenced point and references it to a country. The offshore wind farms were assigned to the closest NUTS 2 area of the country it was referenced too.

The dataset has three categories of hydropower technologies: run-of-river hydropower, reservoir-based hydropower and pumped storage hydropower obtained from the Joint Research Council (JRC) hydropower plants database [39]. The NUTS 2 areas were assigned the cumulative installed capacity of the different hydropower capacities. The reservoir-based and pumped storage hydropower plants' cumulative storage capacity within the NUTS 2 areas was also calculated and added to the dataset. In the instances where the storage capacity was not given, it was assumed that the plant had a reservoir that can store the water needed to operate the plant at nominal capacity for six hours.

The availability of onshore wind and solar were assigned to the dataset as hourly capacity factor values extracted for each NUTS 2 area from renewables.ninja platform [20]. The capacity factors for offshore wind at the national level was taken from the same data platform. The country-wide daily inflow from 46 defines the capacity factor of the hydropower plants provides the historical daily inflow in 30 European countries between 2003 to 2012. The hydropower plants' capacity factors were calculated by dividing the daily inflow values by the sum of installed hydropower capacity within the country.

The data variables that define the area available for renewable energy technologies are rooftop solar PV area, ground-mounted solar PV area, onshore wind area and offshore wind area. For the NUTS 2 areas in the EU-27 countries and the United Kingdom, the ENSPRESO database was used to assign the data variables of rooftop and ground-mounted solar PV area [40] and the data variables for onshore and offshore wind area [44]. The areas classified in the EU-wide low restrictions with 400 m setback distance scenario were selected to define the onshore wind areas. The onshore wind areas in NUTS2 areas of Switzerland were calculated from the wind energy potential areas raster provided by the swiss energy ministry [45]. The onshore wind areas in NUTS 2 areas in Norway were calculated from the wind energy potential areas raster of the hotmaps project [22].

The areas classified within the EU-wide low restrictions with water depth 0 —30 m and water depth 30 —60 m scenario were selected from 43 to determine the offshore wind area. Except for the NUTS 2 areas in Norway, the portion of the offshore wind areas assigned to a NUTS 2 area is proportional to their share of their respective country's total coastline. The NUTS 2 areas of Norway are assigned offshore wind areas according to their proximity to the offshore wind areas listed in the Norwegian offshore wind strategic environmental assessment report[46].

The rooftop and ground-mounted solar PV areas in Switzerland and Norway were calculated using the Open Street Map building footprint data [43] and literature values. For Switzerland, the rooftop solar PV area in a NUTS 2 area was proportional to their share of the total building footprint area in Switzerland multiplied by a total available rooftop area of 267 km^2 and a rooftop suitability factor of 0.564 provided by Walch et al. in [41]. For Norway, the rooftop solar PV area was calculated by multiplying the total building footprint area within the NUTS 2 areas with an rooftop area suitability factor of 0.49 calculated by Bódis et al. in [42]. The ground-mounted solar PV area in Norway and Switzerland was calculated using the ratio of the ground-mounted solar PV area to the rooftop solar PV of Sweden and Austria, respectively. These ratios are 176:1 and 144:1, respectively.

Derived variables

Hourly electrical load profiles for European countries are only available at the country level from the European Network of Transmission System Operators (ENTSO-E) transparency platform 47. The load profiles are given at NUTS 0 spatial resolution in the Areas dataset. While building the Regions dataset, the spatial resolution of the load profiles is first reduced to NUTS 2 level before they are aggregated to the spatial resolution of the defined region. This process is discussed in more detail in the Regions dataset subsection.

Using a bottom-up approach, the heat demand profiles $D_{a,t}$ are generated for each NUST 2 area a and time step t using the following equation:

$$D_{a,t} = D_a \cdot \sum_{s,e} [\sigma_{a,s,e} \cdot d_{a,s,e,t}] \tag{C.1}$$

The bottom-up approach classifies the heat demand in two end-use categories e and two sectors s. The end-use categories are space heating and domestic hot water heating. The sectors are the tertiary and domestic sector. Both end-use categories of each sector have a share factor $\sigma_{a,s,e}$ and a normalised hourly profile $d_{a,s,e,t}$ with time steps t. The share factor gives the percentage contribution of an end-use category of a sector to the total space and water heating demand. These share factors are country-specific and are obtained from the hotmaps repository [22]. The hotmaps repository does not provide share factor values for Norway and Switzerland [22]. Therefore, the share factors for Sweden and Luxemburg were used respectively instead. The normalised profiles are generated at the national level using generic profiles for space heating and water heating obtained from the hotmaps repository [22]. The generic profiles for space heating are country, season and temperature-dependent whereas, the generic profiles of hot water heating only vary according to the day of the week and the season. The normalised space heating profiles are defined using the temperature data in the dataset. NUTS 2 areas within the same country are assigned the same normalised space heating demand profile. The heat demand volume da for space heating and hot water heating is calculated from a rasterised map generated by the hotmaps project [22]. The map depicts the estimated final energy demand for space and water heating on each hectare for EU28, Norway, Iceland and Switzerland for 2015.

The temperature variables from the dataset are used to calculate the hourly efficiency factors of the heat pumps.

The following quadratic regression equation, presented by Ruhnau et al. [49], is used to determine the coefficient of performance $COP_{t,a}$ of the air-source heat pumps:

$$COP_{t,a} = 6.08 - 0.09 \cdot \Delta T_{t,a} + 0.0005 \cdot \Delta T_{t,a}^2 \tag{C.2}$$

Where $\Delta T_{t,a}$ is the temperature difference between the heat sink temperature and the ambient air temperature. The heat sink temperature is assumed to be a constant value of 50°C. As suggested by Ruhnau et al. the calculated $\Delta COP_{t,a}$ is adjusted for real-work effects using a correction factor of 0.85.

Regions dataset. Spatial resolution reduction is often used to reduce the computational demand of solving energy system optimisation problems. Depending on the research question or study focus, the data can be aggregated into regions to reduce the spatial resolution of the dataset. A common spatial resolution reduction method used by energy system modellers is to aggregate the spatial data according to political or administrative boundaries. European countries are classified according to multiple NUTS levels. The political regions method can thus group areas according to the NUTS level specified. For example, the spatial resolution of the data for Germany would reduce from 38 government regions of the NUTS 2 areas to the 16 states of NUTS 1. The spatial resolution could also be further reduced to a national level by aggregating the NUTS 2 area data to the NUTS 0 level. The number of NUTS areas at different levels is dependent on the European country. There are spatial resolution reduction methods that group areas according to the heterogeneity of spatial attributes of NUTS 2 areas. The max-p regions method for example, presented by Fleischer [12], groups areas into regions that are similar in population; wind and solar resource potential; and pumped-hydro storage capacity. The max-p regions method uses the max-p-regions algorithm, introduced by Duque et al. in [50].

Once the regions are defined, the variables of the NUTS 2 areas are aggregated to have variables that represent the regions. The resulting spatially aggregated dataset, hereafter referred to as the Regions dataset, is used to store and organise the variables generated after data aggregation. In the Regions dataset, the NUTS 2 reference key is replaced by the Regions reference key. The Regions reference key is composed of the NUTS 2 reference keys of the NUTS 2 area within the regions created. The geometry of the NUTS 2 areas attributed to the same region are joined to form the geometry of the regions.

As mentioned in the derived variables section, the spatial resolution of the electrical power profile in the Areas dataset is at the NUTS 0 level. Therefore the power profiles need to be disaggregated to NUTS 2 area spatial resolution before they can be aggregated to the specified regions spatial resolution. Population and Gross Domestic Product (GDP) are commonly used as a proxy to determine the distribution of electrical demand [15, 51, 52]. Robinius et al., presents a method to disaggregate electricity demand at sub-national levels, but as this method is determined using data for Germany, it is not applicable to all European countries. The chosen proxy to disaggregate the hourly load profiles in the presented case studies is population. In this proposed data processing approach, the NUTS 2 areas assume a share of the load profiles of their respective country. The proportion of the share is calculated by multiplying the country level power profile with the NUTS 2 area-specific weighing factor. In the case studies presented in this paper, population is used to calculate the weighting factors used to disaggregate the power profiles. The weighting factor of a NUTS 2 area is the share of the population in that area in relation to the population of the NUTS 2 area respective country. This approach could be improved as higher spatial resolved data for power profiles for European countries become available. The offshore wind capacity factors are also at NUTS 0 spatial resolution, similar to power profiles. Therefore, offshore wind capacity factors are also disaggregated to NUTS 2 spatial resolution before aggregating them to build the Regions dataset. The capacity factors for offshore wind at NUTS 2 are assumed to be the same as the respective country-level capacity factors.

When building the Regions dataset, the capacity factors of the variable renewable technologies are multiplied by a weighting factor before they are summed. The proxy variable used to determine the weighting factor is technology-specific. The weighting factor is the share of the proxy variable relative to the proxy variable's sum within a region. The technologies and their respective proxy variables used to calculate the weighting factors are given in Table C.2.

All other variables do not represent mean values and are summed without weighting factors.

TABLE C.2: Proxy variables used to aggregate the capacity factors in the case studies.

Data variable	Proxy variable
Solar capacity factor	Sum of Rooftop solar PV area and ground-mounted solar PV
Wind capacity factor	Onshore wind area
Offshore wind capacity factor	Offshore wind area
Hydropower capacity factor	Hydropower plants dispatch
Air-source capacity factor	Population

Model formulation. There are some additional items needed in conjunction with a Regions dataset to formulate an energy system model. The first item is an energy system framework. There is a selection of open-source energy modelling frameworks that can be used. The selection of the framework depends on the focus of the study and the preference of the modeller. As the Regions database is generated using the python programming language, it can be integrated well into a python-based modelling framework.

Together with some additional items, the Regions dataset can then be used to formulate energy system models. One essential item is the techno-economic parameters. The techno-economic parameters will depend on the scenarios being investigated by the model. The scenarios also dictate certain assumptions used in the model.

C.3 Power and heat optimisation model development

The proposed data processing workflow, implemented in the EUropean Sustainable Energy System (EU-SES) modelling tool [53], is used to build power and heat optimisation models to demonstrate the versatility of the data processing approach and the importance of spatial context in energy system modelling. The EU-SES tool uses the calliope framework [54] to formulate the models. The scripts used to generate the datasets, models and the optimisation results of each model can be found on Zenodo [53]. The current version of the EU-SES tool can only automate the construction of an energy system model using the calliope framework. However, as the datasets are separated from models, the datasets can be used as input data in other modelling frameworks such as PyPSA.

The first model is a multi-national model containing ten countries in the NUTS 2 area dataset named the GER NUTS0 model. These ten countries include Germany and nine countries that have a transmission connection with Germany. The NUTS 2 areas spatial data are aggregated according to national jurisdiction in the GER

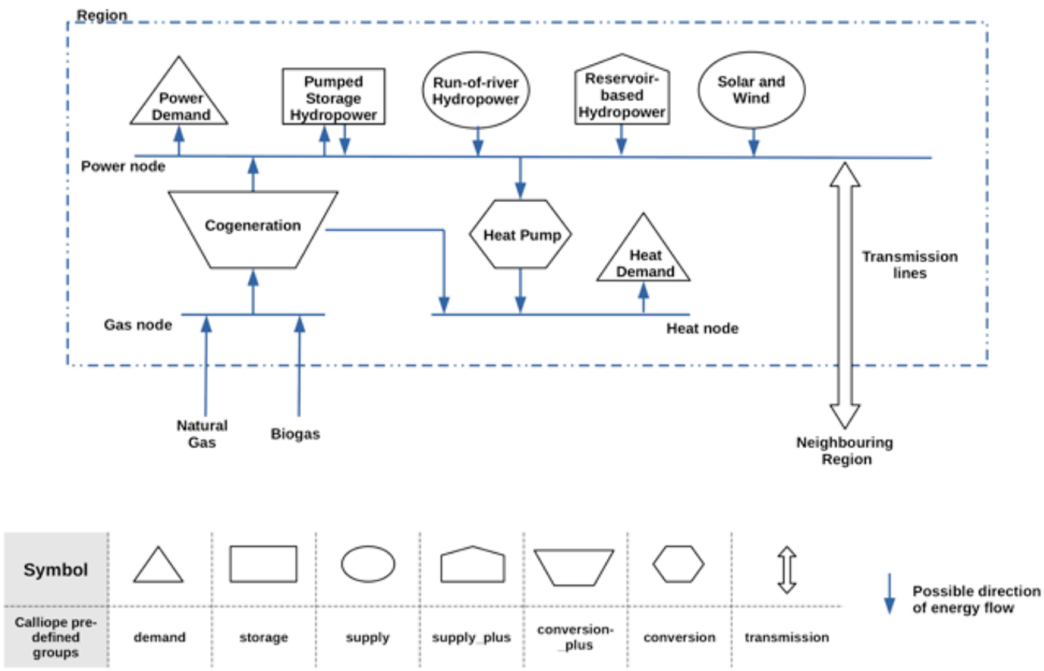


FIGURE C.2: An illustrated description of the model created using the regions dataset and the calliope modelling framework. The table in the figure indicates the predefined technology groups used to describe the different components in the model.

NUTS0 model. The second and third model reduces the spatial scope to include only Germany with no energy import or export from neighbouring countries. The difference between the second and third model is the spatial resolution reduction method used. The second model, named the GER NUTS1 model, are defined according to 16 administrative jurisdictions given by the NUTS 1 level. Whereas the regions in the third model, named the GER MAX-P model, are defined using the max-p regions method to generate nine regions. As illustrated in part a) of Figure C.3, GER NUTS 1 model has more regions and therefore, the regions have, on average, a higher spatial resolution than the regions in the GER MAX-P model. The regions with the highest spatial resolution in the GER NUTS 1 model represent the city-states of Berlin, Hamburg and Bremen.

The reference year selected to create the Areas dataset is 2011. The structure of the models is illustrated in Figure C.2. The examples are modelled using the calliope modelling framework.

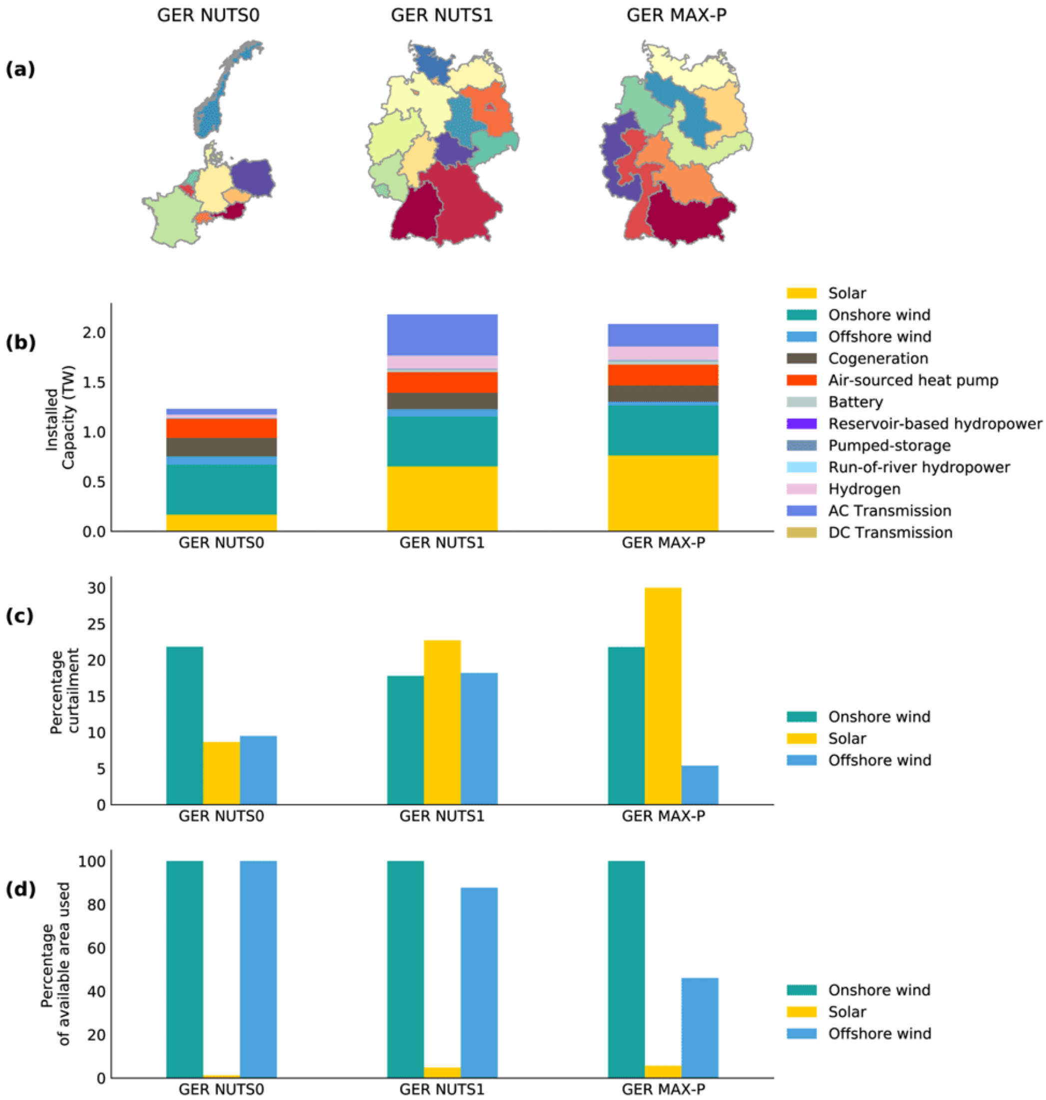


FIGURE C.3: The regions in the GER NUTS0 model, GER NUTS1 model and the GER Max-P model, are illustrated in a). The least-cost optimisation results of the three models for Germany are given in b), c) and d). Plot b) illustrates the optimised installed capacity of the technologies. The curtailment rate of solar PV, onshore and offshore wind generation is given in percentage value in c). Plot d) illustrates the optimised percentage of the available area utilised by solar, onshore and offshore wind installations.

The models all share several overarching scenario assumptions. The following key assumptions are made in this scenario:

- The cumulative CO_2 equivalent emission of the optimised model is limited to 20 % of the 1990 CO_2 equivalent emission of the countries in the model;
- The cumulative biogas available to the cogeneration plants is 420 PJ which was estimated to be a projected value for 2020 presented by Scarlat et al. in [55];
- All power plant capacities classified as biomass, gas and cogeneration in the Regions dataset are summed under the classification cogeneration;
- Power plants classified as nuclear, coal, oil, other, waste and geothermal in the Regions dataset are not available in the model;
- The storage level of all storage capacities is assumed to be full during the first and last instance of the optimisation;
- The hydropower power plants, existing installed wind and solar capacities must be adopted in the optimised model;
- The solar and onshore wind capacity density is assumed to be 170 MW/km^2 and 5 MW/km^2, respectively, adopted from Ruiz et al. [23];
- Offshore wind installations have a capacity density of 5.36 MW/km^2 adopted from Hundleby and Freeman [56];
- Regions are considered "copper plates", meaning that within the regions there are no constraints in energy transfer.

The power exchange between regions is possible and is constrained by the net transfer capacity and the efficiency of the power lines. There are two power transfer mediums in the model. The first is the high voltage alternative current (HVAC) transmission lines between regions that share a border. The HVAC has a set rated capacity of 2 GW. The other power transfer mediums are direct current high voltage interconnectors installed between regions. The list of interconnectors and their respective rated transfer capacity is taken from the installed and planned DC links listed by Fleischer et al. in [12]. Losses are not considered in the interconnectors and HVAC transmission lines.

The power and heat optimisation model's objective function is to minimise the investment cost and dispatch cost of the model for one year and at a three-hour resolution. The optimisation models assume perfect foresight, and the power and heat demand is inelastic. A discount rate of 7% is assumed to calculate the annualised cost of the investments. The model uses techno-economic parameters projected for the year 2030, documented as Extended data [53]. The techno-economic parameters for the generation and storage technologies were adopted from values presented by Moles et al. [57] and by Jülch [58], respectively.

The cumulative CO_2 emission constraint ensures that the models have high solar and wind penetration levels. This emission constraint aligns with the roadmap presented in 2011 by the European Commission that aims to reduce 80% of the EU CO_2 emission by 2050. In 2019 the EU commission revised the CO_2 emission target for 2050 to a net-zero emission target [59]. Therefore the 80% reduction target could represent a snapshot along the net-zero pathway.

C.4 Results and discussions

The optimisation results of the three models are compared in Figure C.3. The results in Figure C.3 show that the GER NUTS 0 model has the lowest installed capacity of solar PV. This is despite the fact that Germany is represented at a lower resolution in the GER NUTS 0 model than the two other models and does not have the opportunity to maximise the use of good solar sites within Germany. This relatively low solar PV installed capacity of the GER NUTS0 model can be explained by the fact that the GER NUTS0 model has a greater spatial scope that the two other models. This additional benefit in spatial scope allows the GER NUTS0 model to maximise the use of resources available in neighbouring countries to Germany, such as hydropower-based energy storage capacities in Norway, Switzerland and Austria. These storage capacities can help minimise the curtailment rate of the solar PV installations, as illustrated in part c) of Figure C.3. The lower-cost hydropower storage capacities in neighbouring countries can also explain why Germany in GER NUTS0 model invest less in expensive hydrogen storage in comparison to the two other models. These apparent differences between the GER NUTS0 model and the models with a different spatial scope document the importance of spatial context in energy system modelling. Part d) of Figure C.3 illustrates that more than half of the available onshore area in Germany is used for deploying onshore wind in all three models.

Next, the optimisation results of the two models with the same spatial scope, the

GER NUTS 1 model and the GER MAX-P model, are presented and discussed. The optimised transmission capacity of the GER NUTS 1 model is significantly greater than that of the GER MAX-P model, as can be seen in part b) of Figure C.3. The fact that the GER NUTS 1 model has more regions, it can have more transmission lines, and therefore it can also have a higher installed transmission capacity value than the GER MAX-P model. As shown in part c) of Figure C.3, the optimised GER NUTS 1 model has a slightly higher percentage in curtailment for solar PV and onshore wind, which could be a consequence of more transmission capacity bottlenecks between regions. The differences between the two models that have the same spatial scope but constructed using two different spatial resolution reduction methods demonstrate the importance of spatial context in energy system modelling.

In the following paragraph some reflections are made on the proposed data processing approach. Firstly, this data processing demonstrates that it is possible to automate the construction of sector-coupled energy system models for European countries using exiting web-hosted datasets. Secondly, there are certain data gaps that influence the data processing approach. The first data gap is the lack of power profiles at sub-national spatial resolution. Due to this data gap the power of profiles of a NUTS 2 area is simply assumed to be a portion of the country-level power profile. The portion of the profile is calculated using population data. This method of disaggregating power profiles does not consider certain differentiations between NUTS 2 areas other than population that influence the power profile such as energy intensive industries in areas with low population. To mitigate this limitation this issue certain models also use GDP when disaggregate power profiles. There are also some data gaps in relation to hydropower plants. These data gaps are power plant specific inflow data and storage capacity of power plants with storage reservoirs. Similarly a limited amount of research has been conducted on the impact of spatial aggregation methods on data products particularly on the impact on capacity factors of variable renewable energy technologies and demand profiles.

C.5 Conclusion

A novel data processing workflow that maximises the use of the web-hosted validated pre-processed input data to build energy system models is presented. The proposed data processing workflow has a two-step process. The first step organises and standardises the pre-processed input data into a dataset called the Areas dataset. In the second step, the spatial data in the Areas dataset is aggregated according to regions and standardised into a Regions dataset. With the addition of techno-economic parameters and a modelling framework, the Regions dataset can be used to build power and heat models. The data processing approach is not integrated into any specific energy modelling framework, giving the modeller the flexibility to create a power and heat model using the modelling framework best suited for the research question. The proposed approach also provides a baseline that can be extended upon to include other energy sectors such as industry and transport. The proposed workflow is used to build three power and heat optimisation models. The three optimisation models' result demonstrates the importance of how the spatial scope and the method used for spatial resolution reduction can impact the optimisation result.

Data availability

Underlying data

Open Science Framework: A data processing approach with built-in spatial resolution reduction methods to construct energy system models.
https://doi.org/10.17605/OSF.IO/JHMXN [53].

This project contains the following underlying data:

- Optimisation results (optimisation results of generated models saved as nc format files)

- Dataset (An areas dataset generated using the proposed data processing saved as nc format file)

Extended data

Open Science Framework: A data processing approach with built-in spatial resolution reduction methods to construct energy system models.
https://doi.org/10.17605/OSF.IO/JHMXN [53].

This project contains the following extended data in Extended data figures and tables.docx:

- Appendix 2 —Techno-economic parameters of storage technologies used the example models.

- Appendix 2 —Techno-economic parameters of storage technologies used the example models.

- Appendix 3 —Percentage difference in installed capacity of GER NUTS1 model without rooftop solar PV in reference to GER NUTS1 model with rooftop solar PV for all NUTS 1 administrative regions of Germany.

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Software availability

- Source code available from:
  https://github.com/ENSYSTRA/EU-SES/tree/v1.3

- Archived source code at time of publication:
  hhttps://doi.org/10.5281/zenodo.5834185 [53].

- License: Apache License 2.0 license.

Acknowledgements

The author would like to thank Md. Nasimul Islam Maruf, André Harewood, Abhinav Bhaskar and Annabelle Constance for their comments and suggestions on the paper. The author is thankful to the ENSYSTRA and ZNES team for support and guidance.

References

[1] Maruf, M. N. I. "Sector Coupling in the North Sea Region—A Review on the Energy System Modelling Perspective". en. In: *Energies* 12.22 (Nov. 2019), p. 4298.

[2] Pfenninger, S., Hirth, L., Schlecht, I., Schmid, E., Wiese, F., Brown, T., Davis, C., Gidden, M., Heinrichs, H., Heuberger, C., Hilpert, S., Krien, U., Matke, C., Nebel, A., Morrison, R., Müller, B., Pleßmann, G., Reeg, M., Richstein, J. C., Shivakumar, A., Staffell, I., Tröndle, T., and Wingenbach, C. "Opening the black box of energy modelling: Strategies and lessons learned". en. In: 19 (Jan. 2018), pp. 63–71.

[3] Pfenninger, S., DeCarolis, J., Hirth, L., Quoilin, S., and Staffell, I. "The importance of open data and software: Is energy research lagging behind?" en. In: *Energy Policy* 101 (Feb. 2017), pp. 211–215.

[4] Nahmmacher, P., Schmid, E., Hirth, L., and Knopf, B. "Carpe Diem: A Novel Approach to Select Representative Days for Long-Term Power System Models with High Shares of Renewable Energy Sources". en. In: *SSRN Electronic Journal* (2014).

[5] Pfenninger, S. "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". en. In: *Applied Energy* 197 (July 2017), pp. 1–13.

[6] Poncelet, K., Delarue, E., Duerinck, J., and Six, D. "Impact of Temporal and Operational Detail in Energy-System Planning Models". en. In: (), p. 18.

[7] Kotzur, L., Markewitz, P., Robinius, M., and Stolten, D. "Impact of different time series aggregation methods on optimal energy system design". In: *Renewable Energy* 117 (Mar. 2018). arXiv: 1708.00420, pp. 474–487.

[8] Heuberger, C. F., Rubin, E. S., Staffell, I., Shah, N., and Mac Dowell, N. "Power capacity expansion planning considering endogenous technology cost learning". en. In: *Applied Energy* 204 (Oct. 2017), pp. 831–845.

[9] Frew, B. A. and Jacobson, M. Z. “Temporal and spatial tradeoffs in power system modeling with assumptions about storage: An application of the POWER model”. en. In: 117 (Dec. 2016), pp. 198–213.

[10] Krishnan, V. and Cole, W. “Evaluating the value of high spatial resolution in national capacity expansion models using ReEDS”. In: *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, July 2016, pp. 1–5.

[11] Hörsch, J. and Brown, T. “The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios”. In: *2017 14th International Conference on the European Energy Market (EEM)*. ISSN: 2165-4093. June 2017, pp. 1–7.

[12] Fleischer, C. E. “Minimising the effects of spatial scale reduction on power system models”. en. In: *Energy Strategy Reviews* 32 (Nov. 2020), p. 100563.

[13] Simoes, S., Zeyringer, M., Mayr, D., Huld, T., Nijs, W., and Schmidt, J. “Impact of different levels of geographical disaggregation of wind and PV electricity generation in large energy system models: A case study for Austria”. en. In: 105 (May 2017), pp. 183–198.

[14] Siala, K. and Mahfouz, M. Y. “Impact of the choice of regions on energy system models”. en. In: 25 (Aug. 2019), pp. 75–85.

[15] Hörsch, J., Hofmann, F., Schlachtberger, D., and Brown, T. “PyPSA-Eur: An Open Optimisation Model of the European Transmission System”. en. In: 22 (Nov. 2018). arXiv: 1806.01613, pp. 207–215.

[16] Tröndle, T. *timtroendle/possibility-for-electricity-autarky: v1.0*. June 2019.

[17] Andresen, G. B., Søndergaard, A. A., and Greiner, M. “Validation of Danish wind time series from a new global renewable energy atlas for energy system analysis”. en. In: *Energy* 93 (Dec. 2015), pp. 1074–1088.

[18] *tum-ens/pyGRETA: python Generator of REnewable Time series and mAps | Zenodo*.

[19] Mattsson, N., Verendel, V., Hedenus, F., and Reichenberg, L. “An autopilot for energy models – Automatic generation of renewable supply curves, hourly capacity factors and hourly synthetic electricity demand for arbitrary world regions”. en. In: 33 (Jan. 2021), p. 100606.

[20] Pfenninger, S. and Staffell, I. “Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data”. en. In: *Energy* 114 (Nov. 2016), pp. 1251–1265.

[21] Wiese, F., Schlecht, I., Bunke, W.-D., Gerbaulet, C., Hirth, L., Jahn, M., Kunz, F., Lorenz, C., Mühlenpfordt, J., Reimann, J., and Schill, W.-P. “Open Power System Data – Frictionless data for electricity system modelling”. en. In: *Applied Energy* 236 (Feb. 2019), pp. 401–409.

[22] Pezzutto, P. S., Zambotti, S., Croce, S., Zambelli, P., Scaramuzzino, C., Pascuas, R. P., Zubaryeva, A., Exner, D., Müller, A., Klingler, A.-L., Kühnbach, M., Manz, P., Marwitz, S., Steinbach, J., Popovski, E., Kranzl, R. L., and Fritz, S. “D2.3 WP2 Report – Open Data Set for the EU28”. en. In: (2018 2018), p. 158.

[23] Ruiz, P., Nijs, W., Tarvydas, D., Sgobbi, A., Zucker, A., Pilli, R., Jonsson, R., Camia, A., Thiel, C., Hoyer-Klick, C., Dalla Longa, F., Kober, T., Badger, J., Volker, P., Elbersen, B. S., Brosowski, A., and Thrän, D. "ENSPRESO - an open, EU-28 wide, transparent and coherent database of wind, solar and biomass energy potentials". en. In: *Energy Strategy Reviews* 26 (Nov. 2019), p. 100379.

[24] *openego/eTraGo*. original-date: 2017-04-05T12:14:57Z. June 2020.

[25] Hauke Henke. *OSeMBE An open source engagement model*. Tech. rep. D7.3. KTH Royal Institute of Technology, Aug. 2018.

[26] Samsó, R. and Ollé, J. S. *Guiding European Policy toward a low-carbon economy. Modelling sustainable Energy system Development under Environmental And Socioeconomic constraints*. Tech. rep.

[27] Pfenninger, S. and Pickering, B. "Calliope: a multi-scale energy systems modelling framework". In: 3.29 (2018), p. 825.

[28] Dorfner, J., Schönleber, K., Dorfner, M., sonercandas, froehlie, smuellr, dogauzrek, WYAUDI, Leonhard-B, lodersky, yunusozsahin, adeeljsid, Zipperle, T., Herzog, S., kais-siala, and Akca, O. *tum-ens/urbs: urbs v1.0.1*. July 2019.

[29] Hilpert, S., Günther, S., Kaldemeyer, C., Krien, U., Plessmann, G., Wiese, F., and Wingenbach, C. "Addressing Energy System Modelling Challenges: The Contribution of the Open Energy Modelling Framework (oemof)". en. In: (2017).

[30] Howells, M., Rogner, H., Strachan, N., Heaps, C., Huntington, H., Kypreos, S., Hughes, A., Silveira, S., DeCarolis, J., Bazillian, M., and Roehrl, A. "OSeMOSYS: The Open Source Energy Modeling System". en. In: *Energy Policy* 39.10 (Oct. 2011), pp. 5850–5870.

[31] Rossum, G. v., Drake, F. L., and Van Rossum, G. *The Python language reference*. eng. Release 3.0.1 [Repr.] Python documentation manual Guido van Rossum; Fred L. Drake [ed.] ; Pt. 2. OCLC: 837611997. Hampton, NH: Python Software Foundation, 2010.

[32] Hoyer, S. and Hamman, J. J. "xarray: N-D labeled Arrays and Datasets in Python". en. In: *Journal of Open Research Software* 5 (Apr. 2017), p. 10.

[33] Geographic Information System of the COmmission. *GISCO data distribution API - NUTS 2016*.

[34] Felice, M. D. and Kavvadias, K. *ERA-NUTS: meteorological time-series based on C3S ERA5 for European regions (1980-2020)*. Version 1980-2020. Mar. 2021.

[35] European Commission, Joint Research Centre; Columbia University, Center for International Earth Science Information Network. *GHS-POP R2015A - GHS population grid, derived from GPW4, multitemporal (1975, 1990, 2000, 2015). European Commission, Joint Research Centre (JRC) [Dataset]*. en. Type: dataset PID: http://data.europa.eu/89h/jrc-ghsl-ghs_pop_gpw4_globe_r2015a. Jan. 2015.

[36] Schlecht, I. and Simic, M. *Renewable power plants*. Version Number: 2020-08-25. Aug. 2020.

[37] Centro Tecnologico del Mar - Fundación CETMAR. *Emodnet HA WindFarms 20200305*. Type: dataset. June 2014.

[38] Gerbaulet, C., Kunz, F., and Weibezahn, J. *Conventional power plants*. type: dataset. 2016.

[39] *JRC Hydro-power plants database. Contribute to energy-modelling-toolkit/hydro-power-database development by creating an account on GitHub*. original-date: 2019-04-05T13:38:23Z. Apr. 2019.

[40] European Commission, Joint Research Centre. *ENSPRESO - SOLAR - PV and CSP. European Commission, Joint Research Centre (JRC) [Dataset]*. en. type: dataset PID: http://data.europa.eu/89h/18eb348b-1420-46b6-978a-fe0b79e30ad3. May 2019.

[41] Walch, A., Castello, R., Mohajeri, N., and Scartezzini, J.-L. "Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty". en. In: *Applied Energy* 262 (Mar. 2020), p. 114404.

[42] Bódis, K., Kougias, I., Jäger-Waldau, A., Taylor, N., and Szabó, S. "A high-resolution geospatial assessment of the rooftop solar photovoltaic potential in the European Union". en. In: *Renewable and Sustainable Energy Reviews* 114 (Oct. 2019), p. 109309.

[43] OpenStreetMap contributors. *Planet dump retrieved from https://planet.osm.org*. Published: https://www.openstreetmap.org. 2017.

[44] European Commission, Joint Research Centre. *ENSPRESO - WIND - ONSHORE and OFFSHORE. European Commission, Joint Research Centre (JRC) [Dataset]*. en. ß. May 2019.

[45] Bundesamt für Energie. *Konzept Windenergie – Grundlagenkarte des Bundes betreffend die hauptsächlichen Windpotenzialgebiete*. Type: dataset. June 2017.

[46] Norwegian water resources and energy directorate (NVE). *Offshore wind power in Norway Strategic environmental assessment– English summary*. Tech. rep. 2013.

[47] Kies, A., Kabitri Chattopadhyay, Elke Lorenz, Lüder von Bremen, and Detlev Heinemann. *Simulation of renewable feed-in for power system*. Tech. rep. Institute of Physics, University of Oldenburg, May 2016.

[48] ENTSO-E. *Detailed Data Descriptions*. European Network of Transmission System Operators for Electricity. 2014.

[49] Ruhnau, O., Hirth, L., and Praktiknjo, A. "Time series of heat demand and heat pump efficiency for energy system modeling". In: *Sci Data* 6.1 (Oct. 2019), p. 189.

[50] Duque, J. C., Anselin, L., and Rey, S. J. "THE MAX-P-REGIONS PROBLEM*". en. In: *Journal of Regional Science* 52.3 (Aug. 2012), pp. 397–419.

[51] Hülk, L., Wienholt, L., Cußmann, I., Müller, U. P., Matke, C., and Kötter, E. "Allocation of annual electricity consumption and power generation capacities across multiple voltage levels in a high spatial resolution". en. In: *International Journal of Sustainable Energy Planning and Management* (Sept. 2017). Artwork Size: 79-92 Pages Publisher: International Journal of Sustainable Energy Planning and Management, 79–92 Pages.

[52] Robinius, M., Stein, F., Schwane, A., and Stolten, D. "A Top-Down Spatially Resolved Electrical Load Model". In: *Energies* 2017 (Mar. 2017).

[53] Fleischer, C. *ENSYSTRA/EU-SES: v1.3*. Version v1.3. Jan. 2022.

[54] Tröndle, T. *euro-calliope: pre-built models*. en. Version Number: 1.0 Type: dataset. July 2020.

[55] Scarlat, N., Dallemand, J.-F., and Fahl, F. "Biogas: Developments and perspectives in Europe". en. In: *Renewable Energy* 129 (Dec. 2018), pp. 457–472.

[56] Giles Hundleby and Kate Freeman. *Unleashing Europe's offshore wind potential - A new resource assessment*. Tech. rep. June 2017.

[57] Moles, C., Sigfusson, B., Spisto, A., Vallei, M., Weidner, E., Giuntoli, J., Jäger-Waldau, A., Lacal Arántegui, R., Lazarou, S., Magagna, D., Carlsson, J., Perez Fortes, M. d. M., Marco, G. d., Jakubcionis, M., European Commission, Joint Research Centre, Institute for Energy and Transport, and SERTIS. *Energy Technology Reference Indicator (ETRI) projections for 2010-2050.* en. OCLC: 1056430159. 2014.

[58] Jülch, V. "Comparison of electricity storage options using levelized cost of storage (LCOS) method". en. In: *Applied Energy* 183 (Dec. 2016), pp. 1594–1606.

[59] EUROPEAN COMMISSION, ed. *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS The European Green Deal*. COM(2019) 640 final. Dec. 2019.