NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

Clustering methods to find representative days for modelling the Portuguese electricity system

João Vasco da Silva Palmeiro

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

CLUSTERING METHODS TO FIND REPRESENTATIVE DAYS FOR MODELLING THE PORTUGUESE ELECTRICITY SYSTEM

by

João Vasco da Silva Palmeiro

Dissertation presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Advisor: Ian James Scott, PhD

July 2021

ACKNOWLEDGEMENTS

Everyone that has, at some point, been a part of my life has, in one way or another, influenced the way I think, thus been part of the process that led to this thesis, and for that I am glad to thank you. Teachers and Professors have nurtured my pursuit of knowledge since I was a little kid, leading me to achieve what I have so far. In particular, I would like to thank Dr Ian Scott, who not only laid out the grounds for this work, but also guided me through this endeavour with consistent dedication, patience, and flexibility, right to the end of it. I also thank Dr Flávio Pinheiro for helping jump start this project. I would also like to acknowledge the contribution of my former colleagues at EU ACER, who have taken my interest in the Energy Markets to a whole new level, and never left any of my ceaseless questions unanswered.

To my friends and colleagues, thank you for all the time together, knowledge shared, and advice given. You have made it much easier for me to take this journey. I have to especially thank Luís Ferreira, for not giving up on me when I needed his help debugging code.

Perhaps most importantly, I would like to thank my family, and my partner in particular, Maria João, for always believing I would see this project to the end, even when it seemed to be so far from it, and supporting me the whole way through in a way only they possibly could.

ABSTRACT

Power system modelling affects decisions on over \$450 billion worth of assets world-wide each year. While complex and computationally demanding models, when properly simplified a balance between accuracy and simulation time can be achieved. Solutions and results for this thoroughly studied problem tend to be rather case-specific, and the Portuguese system presents challenges that make existing approaches insufficient. To better understand this system and how its peculiarities can be used to reduce its modelling complexity, a model of the Portuguese electricity system using PLEXOS software was developed and used to test the impact of different clustering techniques on the model's output results. We show that including natural hydro inflow in the clustering to find representative days for a system where hydro generation plays such a large role can improve model output accuracy. This is typically ignored in the literature. Additionally, we demonstrate that using data disregarding daylight saving time changes can have an impact on results. Finally, we indicate that intraday downsampling might have limited effect on modelling accuracy, and open the way for future work on weighting clustering input dimensions differently to improve accuracy of representative days.

KEYWORDS

Clustering; Energy Economics; Power System Modelling; Renewable Energy Sources; Representative days; Time Series

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

CCGT	Combined Cycle Gas Turbine
CET	Central European Time
CGTEP	Combined Generation and Transmission Expansion Planning
CO_2	Carbon Dioxide
CUC	Clustered Unit Commitment
DGEG	*Direção Geral de Energia e Geologia*
DST	Daylight Saving Time
DTW	Dynamic Time Warping
ENTSO-E	European Network of Transmission System Operators for Electricity
ETS	Emissions Trading System
EU	European Union
EV	Electric Vehicle
FC	Fixed Cost
GEP	Generation Expansion Planning
GJ	Gigajoule
GWh	Gigawatt-hour
IAM	Integrated Assessment Models
IEA	International Energy Agency
IRES	Intermittent Renewable Energy Sources
MIBEL	*Mercado Ibérico de Electricidade*
MW	Megawatt
MWh	Megawatt-hour
NEM	National Electricity Market
NRMSD	Normalised Root Mean Square Deviation
OCGT	Open Cycle Gas Turbine
PV	Photovoltaic

RES	Renewable Energy Sources
SRMC	Short-Run Marginal Cost
TEP	Transmission Expansion Planning
TP	Transparency Platform
UTC	Coordinated Universal Time
VoLL	Value of Lost Load
V2G	Vehicle-to-Grid

1. INTRODUCTION

The supply of electricity needs to meet demand for electricity almost precisely at every moment in time: that is the major uniqueness of Electricity Markets in comparison to all other markets. We need to fully understand this physical phenomenon to comprehend today's market design. Furthermore, electricity is considered an essential commodity to the welfare of modern society, meaning that shortages and blackouts can have grave negative impacts.

As the electricity sector evolved through the past century, so did the ideas around the role of the State in the market for electricity (Figueira, 2018; Hannah, 1979; Matos, Mendes, Faria, & Cruz, 2004). In the second half of the 20th Century the paradigm changed from the State being at the same time stakeholder, generator, transmitter, distributor, retailer, and (self)regulator to a more decentralised, dynamic, and competitive market with independent regulators (Fortunato et al., 2008).

The Single European Act, signed in 1986, set the goals of liberalization of the parts of the energy markets that could be efficiently liberalised, and of integrating the separate European markets into a single market. Only the first goal has been mostly accomplished so far, as since July 2007 all consumers in the European Union (EU)[1] are free to choose their energy provider for both electricity and natural gas, making it still relevant to study each market individually (Vasconcelos, 2019).

Because of the markets' size, the peculiarities of each of the islands, the lack of granular data for the islands, and the markets being physically separated, this study will only encompass the Portuguese mainland, referring only to the mainland when the Portuguese electricity market, system, and grid are referenced.

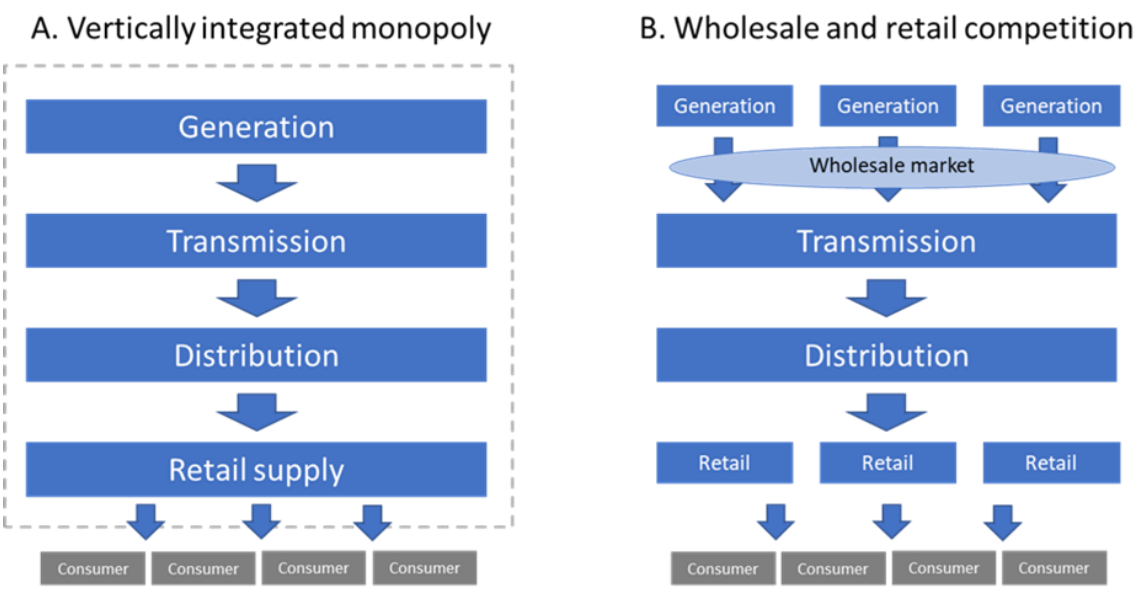

Figure 1 – A. Diagram exemplifying a vertically integrated electricity market monopoly.
B. Diagram exemplifying a wholesale and retail competition electricity market.

[1] Except for some islands and new EU Member States.

Working up to downstream on the electricity market as it is shaped in Portugal nowadays (Figure 1.B) we start with the actual energy generation or production. The unbundling of the incumbent company (from Figure 1.A to Figure 1.B) and gradual liberalization of the sector was started in 1994 (Vasconcelos, 2019). The electricity generation market is today a relatively free market, where any company can participate. However, as grid capacity is limited, new generation capacity cannot be connected to the grid at will. There are several ways to be granted access to the grid, with the most commonly used system being public auctions ran by the Transmission System Operator, *Redes Energéticas* Nacionais, with the supervision of the National Regulatory Authority, *Entidade Reguladora dos Serviços Energéticos,* emphasising the need for accurate system modelling in order to achieve optimal capacity installation and generation mix[2].

In Energy Economics studies, similar electricity generation facilities are usually grouped into generation types. Different generation types have distinct characteristics and so are used to meet different ends. Since supply needs to meet demand at all points in time, there has to be enough installed capacity of dispatchable electricity generation to cover peak net demand[3]. There are only a few hours in a year when demand reaches certain peaks, meaning that to meet these peaks there would need to be power plants built solely to run for a few hours a year. The decision to build these plants is based on the comparison of the Value of Lost Load (VoLL)[4] with the average cost of these powerplants for the few hours they would run.

Economically, the decision of which powerplant(s) to build depends on the combination of their fixed cost (FC) and short-run marginal cost (SRMC). Powerplants with a large FC and low SRMC will be built to meet baseload demand, i.e., to run for most hours in the year, while powerplants with low FC and high SRMC will be built to meet peak demand, running for less time.

There are also strategic (e.g., fuel diversification), political (e.g., fuel dependency), environmental (e.g., emissions, noise, visual, landscape) and technology-related (e.g., dispatchability, capacity factor[5]) reasons to choose different generation technologies. Some can be translated into market-based approaches, as the case of emissions with the European Union Emissions Trading System (EU ETS). This cap and trade scheme introduced in the EU in 2013 sets a cap on emissions for each country that are auctioned for the countries' larger polluting agents and can then be traded, forming a market price for emissions (European Comission, 2015). Other factors such as the low-risk high-cost chance of a nuclear disaster, or the dependency on fuel from other countries are more difficult to translate into figures.

[2] Generation mix is the combination of the different electricity generation types used to meet demand at a given time.

[3] Because, by definition, Intermittent Renewable Energy Sources (IRES) cannot be used at will, and it cannot be guaranteed that they will be producing at the moments of peak demand, the only way to ensure there are no blackouts at peak times is to have dispatchable electricity generators that cover demand net of renewable generation and imports at each point in time.

[4] VoLL is usually defined as the value attributed by consumers to unsupplied energy (Glowacki, 2016).

[5] Capacity factor is a measure of the overall usage rate of a powerplant. It is calculated as the actual generation of a powerplant in a given period over the total possible generation for that period (capacity times the number of hours in that period) (Morales Pedraza, 2019).

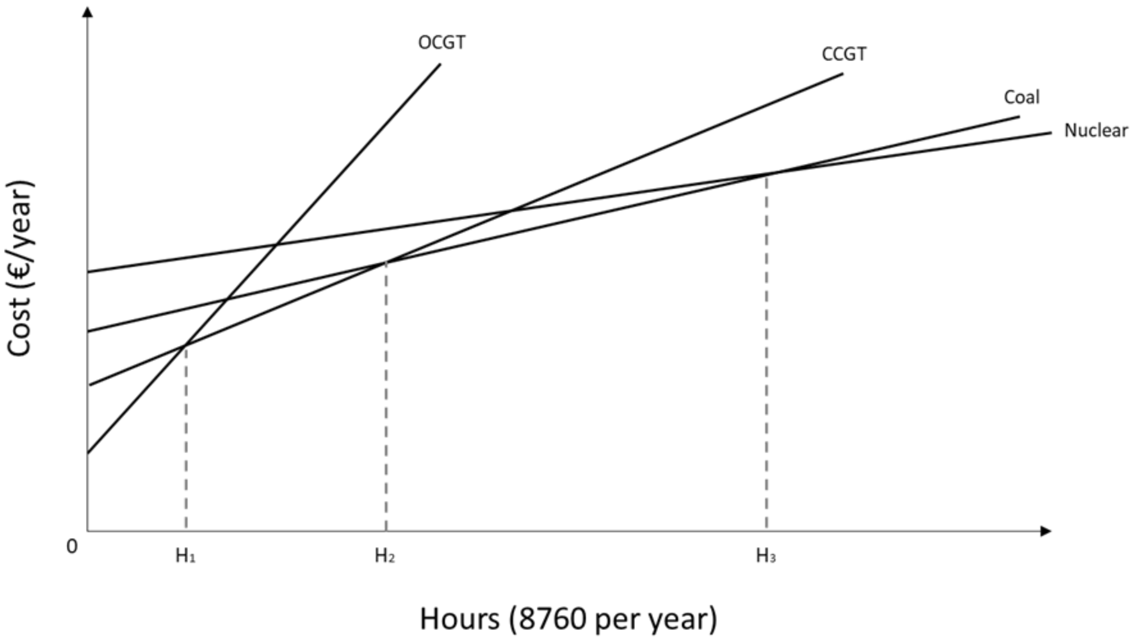

Figure 2 – Theoretical comparison of total cost and running time of fossil fuel and nuclear generators – screening curve.

In Figure 2 we can see how different combinations of FC and SRMC lead to different build decisions depending on the running time predicted for a powerplant. The lowest curve at each point indicates the lowest cost technology for a powerplant that needs to run for that many hours in a year. In this example, if the running time for a certain capacity in a year is predicted to be higher than 0 and up to H1 then an open cycle gas turbine (OCGT) plant would be built because it has the lowest FC even though it has the highest SRMC; if between H1 and H2 then combined cycle gas turbine (CCGT); if between H2 and H3 then a coal plant; and if higher than H3 a nuclear powerplant would be built since it has the lowest SRMC even though the FCs are the highest.

After the build decisions have been settled, there needs to be decided which units will be running at each timeframe (usually hourly). In a perfect competition scenario, the generation mix is formed by arranging the different generators in their merit order by SRMC (subject to technical constraints). Then, for the demand at a given point in time the market price should be the SRMC of the last generator in the merit order to be generating. In Figure 3 we can find an example of such merit order. In this example, and ignoring technical constraints, if demand equals D1, then all generators would be producing at full capacity except for OCGT that would be producing only enough electricity to meet what is left of demand. If in the next hour demand went down to D2, OCGT would stop generating, and CCGT would produce slightly less.

The decision of which generation units to run will also settle the wholesale market price of electricity for that timeframe. In the example from Figure 3, the market price when generating D1 would be P1, the SRMC of OCGT, meaning that OCGT would be breaking even, and the rest of the generators would be making a profit or paying for the investments. If demand went down to D2 the market price would decrease to CCGT's SRMC (P2).

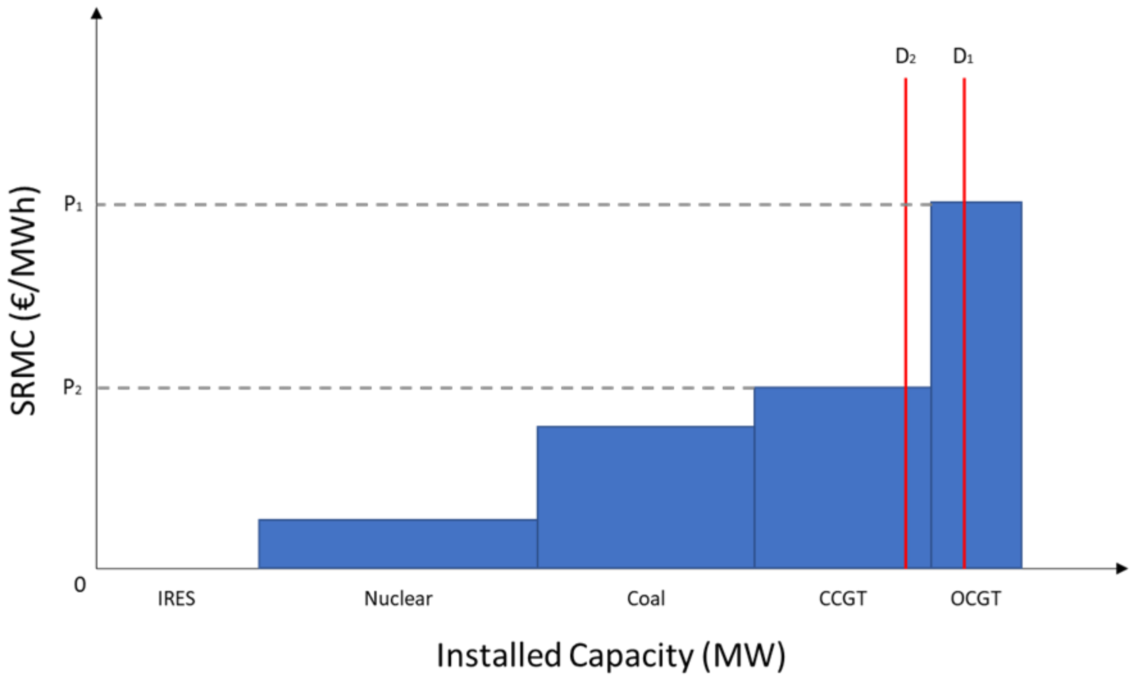

Figure 3 – Example of generation merit order.

In the real markets, there are many more generators and constraints such as ramp up/down costs, minimum generation levels, minimum down times, maintenance stops, subsidies (usually for renewables), and different market timeframes, making the price formation a much more complex process.

This problem also gets less simple when considering renewable energy sources (RES). RES can be dispatchable, meaning that the timing of the electricity production can be controlled, or intermittent (IRES) when generation depends on the natural availability of a resource (although it is usually possible to curtail the generation). The most common examples of IRES are solar, wind and hydro run-of-river, where generation depends on solar radiation, wind and river flow, respectively, and the marginal cost is null. Dispatchable RES can have a set marginal cost, such as in the case of biomass or biogas generators where the marginal cost is linked to the cost of the (renewable) fuel burnt, or they can have an opportunity cost as it happens with hydro reservoir where the water available to generate is limited so it cannot be generating at all times even though the resource was not paid for. In the case of hydro pumped storage the marginal cost is a mix of the opportunity cost (because of the limited amount of water) and the cost of the energy used to pump the water upstream as a way to store energy.

This market model for electricity generation then leads to a liberalised market where different timeframes lead to different prices given by a match between bids and offers. In Portugal, since 2001 the market has been integrated with the Spanish one, forming *Mercado Ibérico de Electricidade* (MIBEL), however the two countries still remain as separate bidding-zones. Even though all electricity sold at MIBEL is the same, green generation can already be traded as a separate product[6].

[6] No consistent and reliable green electricity demand data was found that could be introduced in the model.

The above-mentioned market characteristics need to be considered when modelling an electricity system. This can be done to a greater or lesser extent depending on the detail and complexity needed for the model. These characteristics need to be combined to the particular way each market is either set or evolving towards.

The way electricity is generated and consumed is also in constant evolution. In Portugal, between 2016 and 2020, there was a 74% increase in hydro pumped storage generation capacity, while hydro reservoir remained stable and hydro run-of-river actually decreased over 4% in generation capacity (ENTSO-E, 2021). This comes to aid the problem of generation time adequacy because pump storage is to this date the most efficient and green way to store energy at a large scale, providing better control over electricity generation, and reducing the volatility introduced by rapidly growing IRES, such as solar and wind.

New vehicle-to-grid (V2G) solutions have been presented as a future option for energy storage by taking advantage of storage capacity that would exist anyway but is not being used for lack of storage management. V2G means electric vehicles charging timings being smartly managed, charging off-peak (valley filling) and consuming energy from the batteries during peak demand (peak shaving), providing grid flexibility to minimize demand peaks (Wagner, 2013). V2G solutions are already being tested in Portugal (Energias de Portugal, 2018; SGS, 2021).

Active consumers', commonly called prosumers for being both producers and consumers of electricity, tendency to grow can bring further variability to the power system, negatively affecting grid management (European Comission, 2019; Šajn, 2016; Vasconcelos, 2019).

To this date, V2G and prosumers do not have enough impact in the Portuguese electricity system to have meaningful impact on today's systems, nor is there readily available data on their behaviour in the system. Nevertheless, V2G, prosumers, green energy demand, and cross-zonal trades[7] are all challenges that can be modelled and introduce complexity into the simulation.

According to ENTSO-E (2021), which only includes units with net generation capacity equal to or greater than 1 Megawatt (MW), wind generation capacity in Portugal grew from 4617 MW in 2016 to 5183 MW at the start of 2021 (12%). Even though solar generation installed capacity was much lower than wind's in 2016 at 251 MW, it expanded at a much faster pace to 569 MW, increasing 127% over the same period. The more comprehensive data published by the national entity for energy (DGEG – *Direção Geral de Energia e Geologia*) show a smaller increase of wind (3%) and solar (105%) generation capacity over the same five-year period (Direção Geral de Energia e Geologia, 2021). Regardless of the source the conclusions are similar, wind generation is already a very representative part of Portugal's electricity generation (about a quarter of total installed capacity), and solar generation has been growing significantly and is expected to keep that tendency at least in the near future (Prado, 2020).

[7] Electricity exchanges between different bidding-zones (e.g., between Portugal and Spain).

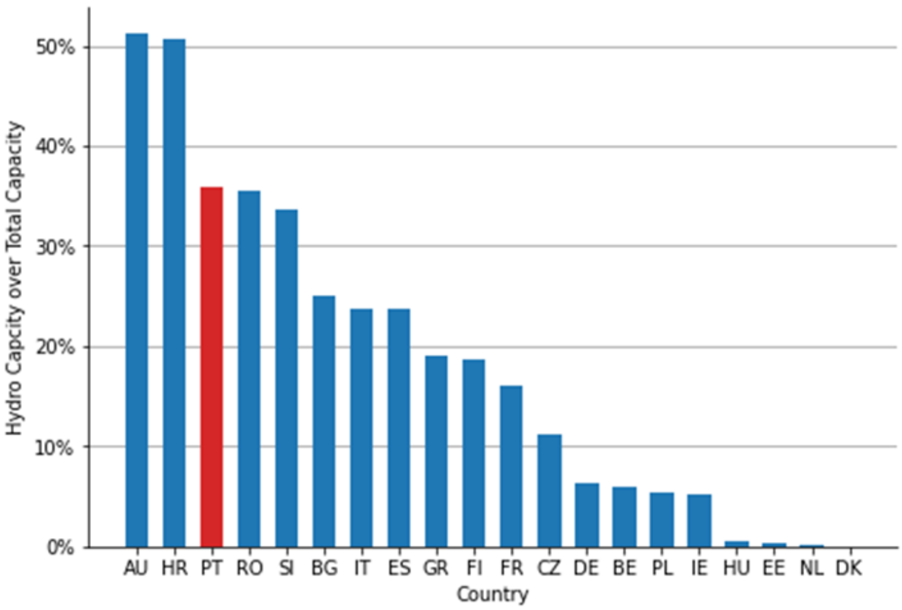

Figure 4 - Hydro generation capacity as a percentage of total installed capacity in European countries in 2021. Source: ENTSOE (2021).

Figure 4 presents the hydro generation capacity as a percentage of total capacity in major EU power systems. It showcases a particularity of the Portuguese power system where hydro generation plays a very important role, representing over one third of total capacity, thus making modelling rather dependent on hydro availability.

The International Energy Agency (2019) reports that in 2018 over $450 billion were invested globally in electricity generation capacity, of which more than half was on RES. These investment decisions affect not only everyone who consumes electricity, but also literally every living being on the planet since it deeply impacts resource usage and pollution. As of 2016 electricity and heating production accounted for more than 40% of global CO_2 emissions (International Energy Agency, 2018).

It is clear then that the electricity sector is an important and changing one. To make decisions in these areas we rely on the use of models. However, these models are large and complex, and we need to reduce complexity where possible. One way to do that is through representative days.

In the remainder of this thesis we review the literature on power system modelling and their purpose, the selection of representative days and the measuring of their accuracy. We then describe how we built such model for the Portuguese electricity system, reduced its complexity, and measured the accuracy of said complexity reduction.

2. LITERATURE REVIEW

Since the creation of the first energy system models by the International Energy Agency (IEA) and the International Institute for Applied Systems Analysis in the 1970s (Pfenninger, 2017) the energy systems have become much more complex and variable. Power system modelling has continued to evolve, allowing build and policy decisions to be reliably data driven.

Power systems are only one category over a multitude of model types, varying in scope and aim, that have been used to different ends (Poncelet, 2018; Scott, 2021):

- Integrated assessment models (IAMs) study long-term interdisciplinary problems of a global scope. Several authors (see e.g., Clarke et al., 2014; Moss et al., 2010) used IAMs to analyse policies for climate change mitigation.
- Energy-economy models are used to study the interaction between an energy and an economic system. These are usually modelled at a national or regional level and with a time scope between 20 and 100 years. Messner & Schrattenholzer (2000) linked a macroeconomic model with a detailed energy supply model to integrate the influence of energy supply costs in macroeconomic production factors optimisation.
- Energy-system planning models have their scope limited to the energy systems in particular, usually modelling the entire chain from extraction to final energy consumption in all major forms for a particular country or region over multiple decades. Götz, Blesl, Fahl, & Voß (2012) used these models to study how to set policy targets to reduce greenhouse gas emission across EU ETS and non-ETS sectors.
- Power-system planning models' scope is restricted to only the power system itself, with the upside of allowing more detailed representations of such complex models. These models can be further categorised:
  - Generation expansion planning (GEP) is used to understand which generation units should be installed or decommissioned to meet expect demand over a planning horizon. GEP usually ignores or greatly simplifies transmission costs.
  - Transmission expansion planning (TEP) aims to minimise transmission costs.
  - Combined generation and transmission expansion planning (CGTEP) takes into account both the need to plan installed capacity and its location, considering the transmission costs associated.

Our model is framed as a power system planning model, in particular market monitoring of unit commitment, which needs to be taken into account by GEP models. For the remainder of this section we will detail how power system models in particular have been used, their complexity reduced, and their accuracy measured.

Krajačić, Duić, & Carvalho (2011) used the H_2RES model to perform system planning and present technical solutions for 100% RES electricity production scenarios in Portugal. The authors showed that a 100% RES solution favours hydro and wind power, with large pump storage hydro facilities to avoid unnecessary rejection of variable renewable generation and smooth net demand curves.

Elliston, MacGill, & Diesendorf (2013) simulated the Australian power system to seek the least costly solutions for supplying the Australian National Electricity Market (NEM) with 100% RES electricity in

2030. The authors found that, depending on the discount rate and the future emission prices, going 100% renewable could be a cheaper solution for the Australian NEM.

Pillai & Bak-Jensen (2010) studied how increasing electric vehicle (EV) loads affects a typical Danish primary distribution network, both with controlled and uncontrolled charging modes. The study concluded that only a 10% (of total cars) integration of uncontrolled charging EVs is feasible, with a much larger integration to be possible with controlled charging. For this study the authors used a model of the power system of the Danish island of Bornholm.

The first two above-mentioned examples emphasise the evermore recurring need to steer focus into RES when modelling power systems, as they are more difficult to model and predict due to intrinsic volatility and at the same time are growing in contribution to the systems. The third highlights the advantages of introducing V2G and prosumers to power system models.

Simulating an electricity system over many years and correctly considering investment decisions, medium-term constraints, and financial incentives can prove computationally difficult, especially when needed to run thousands of times using Monte Carlo methods to estimate marginal costs (Booth, 1972; Mazumdar & Chrzan, 1995). There are 8760 hours in a non-leap year for which inputs and constraints need to be considered, both separately and taking into consideration the previous and following hours' generation profile. A decision to generate at a given time is not independently taken due to opportunity costs, ramp up and ramp down costs, minimum generation levels, minimum up and down time, and other constraints.

In order to account for some of the aforementioned chronological dependency, the most common aggregation for electricity demand data in the literature is representative days (Yeganefar, Amin-Naseri, & Sheikh-El-Eslami, 2020). However, the methods to select these typical days in a way that maintains the essential variability for the models differ considerably (Kotzur, Markewitz, Robinius, & Stolten, 2018). Aggregating hourly year-long profiles into representative days or weeks can considerably reduce this massive computational requirement and, if properly achieved, maintain high accuracy while allowing for more and quicker simulations. Green, Staffell, & Vasilakos (2014) have shown that clustering year-long profiles into 6 to 10 representative days can increase the processing speed by a factor of over 100.

Having established the value of accurately modelling electricity systems, and the need to reduce the complexity and computational cost of such systems, the next step is to consider how this complexity reduction has previously been undertaken.

There are academic studies that look solely into clustering demand data (Hassan, Khosravi, Jaafar, & Raza, 2014), usually net of IRES (Sisternes & Webster, 2013; Yeganefar et al., 2020), studies that cluster demand and wind generation separately (Green et al., 2014), and studies that consider load, wind, and solar generation as IRES become a considerable part of installed capacity (Merrick, 2016; Nahmmacher, Schmid, Hirth, & Knopf, 2016; Poncelet, Hoschle, Delarue, Virag, & Drhaeseleer, 2017). Studies have used other dimensions, usually to different ends.

Several authors (see e.g., Merrick, 2016; Nahmmacher et al., 2016; Poncelet et al., 2017) have all clearly stated that, as IRES become an increasingly important part of energy systems, using net demand to find representative days tends to deeply underestimate of the variability introduced by IRES. These studies emphasize the importance of including IRES along with demand on future works to find representative days for systems with significant IRES penetration. This means that, to select representative days, we should include each series individually rather than diluting them by netting demand. Furthermore, as markets evolve we should investigate which new dimensions can be further included to improve selection.

Pina, Silva, & Ferrão (2011) developed new modelling methodologies using a typical weekday, Saturday, and Sunday for each season for the Portuguese island of São Miguel (Azores). Unlike our study, the aim of their research was to examine modelling techniques rather than clustering techniques, and to validate the study for the peculiar insular electricity system of São Miguel island. To the best of our knowledge, no studies on clustering for modelling only the Portuguese mainland electricity system have been published.

To understand how accurate dimension reduction methods are we can compare and measure model input data and/or model output(s) (Kristiansen, Korpås, & Härtel, 2017). Härtel, Kristiansen, & Korpås (2017) have shown that the most accurate sampling technique when comparing the model input data will not necessarily yield the most accurate model output. On the one hand, assessing accuracy in terms of model input data can present a more generalisable conclusion. However, the more case-specific approach of assessing accuracy in terms of model output can show how clustering actually does impact the end-goal of modelling.

Studies have used both methods to compare sampling techniques applied to power system planning models. Within each method, the variables used to make the comparisons also differ.

For measuring accuracy in terms of model input data Kotzur et al. (2018) used solar irradiation, temperature, electricity load, and wind profile, whereas Nahmmacher et al. (2016) compared the daily profiles of onshore wind, solar photovoltaic (PV), and electricity demand.

Liu, Sioshansi, & Conejo (2018), Kristiansen et al (2017), and Scott, Carvalho, Botterud, & Silva (2019) analysed results using both methods. The first study compared wind, solar, and demand daily duration curves (input data) and investment decisions (model output). Kristiansen et al. (2017) compared load, onshore wind, offshore wind, solar, and hydro (input data) and operational cost performance (model output), showing that the ranking of sampling techniques was not the same with both comparison methods. Scott et al. (2019) analysed the normalised root mean square deviation (NRMSD) of duration curves of demand, wind, solar and ramp (model inputs), and also compared expansion model results.

However, measuring the dimension reduction techniques' accuracy was found most commonly in terms of power system planning model outputs. Green et al. (2014) used electricity cost, carbon intensity, annual output and revenue, and number of plant start-ups and outages. Teichgraeber & Brandt (2019) used problem specific objective functions (battery charge/discharge optimization and gas turbine scheduling). Pfenninger (2017) compared deployed capacity of key technologies and the levelized cost of electricity. Yeganefar et al. (2020) used new capacity added to the generation fleet (long-term planning modelling). Sisternes & Webster (2013) used generators' capacity and commitment. Assessing each model output individually allows a very comprehensive analysis of the results, but it also makes them difficult to process and analyse. Instead, Merrick (2016) used a single comparison metric composed of several model outputs. This solution makes it much simpler to compare results but hides how clustering variations affect different model outputs individually, and is highly dependent on the relevance of the combination of outputs used.

The tendency seems to be putting more focus on assessing if a clustering technique can provide accurate model results. Nevertheless, always bearing in mind the results are more case specific. This emphasises the importance of building an accurate power system model to compare on, that still is general enough to allow extrapolating results.

Previous studies not only vary in the way the model is built and assessed, but also in the clustering techniques used.

Instead of using complex clustering methods, one could consider simply grouping hours with (almost) the same demand, however it has been shown that such does not occur frequently within the same year. According to Merrick (2016), with less than 40 clusters one can capture the vast majority of the variance of a year’s 8760 hours of demand. However, if also including a wind and a solar profile, approximately 1000 hours are required to capture similarly low variance. Even though this is already a considerable downsize from the original 8760 hours, it is not enough of a computation reduction for some studies. Furthermore, this would assume there are no constraints that span across time, i.e., that hours could be treated as chronologically independent. As described in section 1, the models should consider the hourly sequencies. For that reason most studies take into account constraints across time by using representative days (or weeks), assuming chronological independency between days (or weeks). In this case, 20 days out of 365 in a year would be enough to have a low level of variance if only gross demand was to be modelled, rising to 300 days if considering wind and solar profiles. The results are even worse when aggregating weekly, as no two identical weeks were found in Merrick’s study when considering load and availability of wind and solar (Merrick, 2016).

Representative load curves are typical daily curves representing a group of load profiles with analogous demand patterns. Introduced by Balachandra & Chandru (1999), were later used by Green et al. (2014) to demonstrate how complete year-long profiles of the British electricity system can be processed about 60 times faster using only a set of 6 to 10 representative hours and still yield accurate results for estimation of average and marginal cost of electricity. However, their results when modelling rare events such as plant starts, outages and peak requirements were shown to be much less accurate, with clustering results grossly underestimating them. This comes to show that approaches that assume chronological independency become less accurate and thus less relevant with the increase of IRES.

In these studies the k-means algorithm (or a variant of it) was used to cluster the data set as a whole. However, the hardest parts to accurately model are in both ends of the demand spectrum, i.e., the high peaks and the low plunges, due to the nature of marginal costs of the generation mix.

Pineda & Morales (2018) used a variant of k-means by choosing a medoid only after performing all k-means iterations in order to reduce smoothing results.

Liu et al. (2018) and Teichgraeber & Brandt (2019) used dynamic time warping (DTW) distance as a shape-based clustering method to find representative days. Introduced in Sakoe & Chiba (1971, 1978) applied to speech recognition, DTW finds optimal alignment between two sequential sets of data, having been demonstrated to have meaningful applications with time series (Petitjean, Ketterlin, & Gançarski, 2011).

Pfenninger (2017) used various combinations of three different methods to increase the computational tractability of high-resolution planning model of Great Britain's electricity system with 25 years of simulated wind and solar PV generation: downsampling, clustering, and heuristics. This study differs from most others for including a longer than usual dataset, combining methods that tackle different aspects of the problem, and comparing simulations of the same model with different scenarios of share of IRES in the generation mix. Downsampling is the simplest of the three methods, consisting of reducing the resolution of the whole series (e.g., from hourly to 3-hourly). However simple a method, downsampling tended to worsen results when modelling with a high share of renewables, as high variability usually requires high time resolution to be correctly modelled. To cluster, the author used both k-means and hierarchical approaches, describing that the sum of squared error tended to flatten off between 10 and 15 clusters. The third method used, heuristic selection, refers to selecting days or weeks based on pre-defined criteria such as the week containing the maximum and minimum daily average of a time series. When combined with clustering, heuristic selection allows to select extreme days that would otherwise be flattened out even though they can be very important to model. The author concluded that approaches including heuristic methods tended to yield stabler results and can therefore be preferable for models with a high share of IRES.

From this literature we conclude that power system modelling has an important economic impact, and reducing its complexity is a pertinent problem that is becoming increasingly more difficult as markets evolve into more volatile scenarios (e.g., with more IRES). The most widely used approach to this problem is to select representative days rather than hours due to chronological dependencies. To select these days we should cluster all IRES individually and not use net demand, and to assess their accuracy consider the effect on model outputs rather than inputs. However, from the literature it is not clear how a hydro dominated system like Portugal, with the associated additional weather dependant variables, should be modelled.

3. RESEARCH QUESTIONS AND OBJECTIVES

Various studies, namely the ones mentioned in Section 2, have explored multiple methods to reduce the complexity and computational requirements of electricity system modelling, and to evaluate the accuracy of said methods.

This research aims to better understand how techniques to reduce modelling complexity can be efficiently applied to the Portuguese power system in particular. We sought to adapt and combine some of the techniques mentioned in the previous section to the specific characteristics of the Portuguese system. We studied the effects of different clustering techniques using three different methods to compare the accuracy of the clustering, setting out to address the following questions:

> **Q1:** How many representative days are needed to effectively model a year of the Portuguese power system?

The number of representative days needed to accurately model a power system directly affects how much of the modelling costs and computational demand can be reduced. As the number of representative days increases, so should the modelling accuracy. However, this tends to happen at increasingly smaller improvement rates, providing a trade-off between modelling costs and accuracy.

> **Q2:** Is the clustering accuracy affected by using input data that ignore daylight saving time (DST) change (time seasonality)?

This question in particular was not found in any of the literature reviewed. Time series data in UTC (Coordinated Universal Time) ignores daylight saving time changes and is commonly used for time series datasets since it avoids having an hour with missing data and another with two records each year. When using a dataset that considers DST, such as CET (Central European Time), those two hours a year need to be fixed. However, people tend to have a schedule that takes the DST into consideration, meaning that the influences on electricity demand profiles from people's quotidian activities tend to be always at the same CET time, but not UTC. For example, if a factory starts working at 9 a.m. CET, it will provoke an increase in demand at 9 a.m. CET the whole year, but at 9 a.m. UTC half of the year and 10 a.m. UTC the other half. However, as this does not impact the profiles of IRES, the effects of time seasonality on the clustering might change when including IRES series.

> **Q3:** Does the unusual prevalence of hydro generation capacity of the Portuguese system mean that hydro generation data can improve the clustering accuracy?

Even though it is becoming increasingly recurrent to include wind and solar data in the clustering for complexity reduction studies (see section 2), hydro natural inflow tends to be left out of the clustering. Because, contrarily to most systems studied in the literature reviewed, over a third of the installed capacity in Portugal is hydro (see Figure 4), this can be a meaningful dimension to include when clustering for the Portuguese system.

> **Q4:** Should all input data dimensions be given the same weight when clustering?

In section 2 we have described various approaches to clustering representative days for power system modelling, with studies including different model input data in their clustering, and with that having different results. Some dimensions are more volatile between days (see section 4.2) and harder to represent (see section 4.3), but if these are not more important for the model then including them could maximise representativeness of unimportant series. However, incorporating a certain dimension into the clustering does not need to be a zero-sum game. We developed a new approach and studied how weighting differently the various dimensions in the clustering affects the model's accuracy. For example, we are introducing hydro natural inflow. Would this be given the same weight as what we were already considering? Hydro is very important for generation capacity but maybe less volatile. Can we understand how the volatility, correlation with demand, and capacity of that type of generation affect the weighting given to that dimension?

> **Q5:** Do intraday aggregations (downsampling) have a significant impact on the model's accuracy when combined with other techniques?

The number of representative days can only be reduced up to a certain point before accuracy starts dropping drastically. However, a model's complexity can also be reduced with intraday aggregations, such as downsampling. Pfenninger (2017) showed that downsampling can worsen results when clustering with high shares of renewables. However, the major renewables in Portugal are hydro and wind, which present much less intraday variability than for example solar (see section 4.2). Furthermore, we intend to combine downsampling with the previously described weighting technique.

In order to answer these questions we started by collecting, analysing, and preprocessing the data. We then constructed and benchmarked a model of the Portuguese electricity system, and developed both a weighting and a downsampling approach to subsequently cluster the data (with and without these two approaches). Finally, recursively ran the model with the clustered data, extracted and compared the results.

4. METHODOLOGY

In this section we describe the overall methodology used for this study. It starts with collecting the data that is then used to create a working model of the Portuguese electricity system (section 4.1). Then we analysed the model input data to better understand how it behaves in order to efficiently reduce its complexity (section 4.2). Afterwards the data was preprocessed using the techniques developed for this study (section 4.3) before being clustered (section 4.4). Finally, we had to develop metrics to effectively measure the accuracy of the previously applied complexity reduction techniques, comparing model results with the ones of the original model (section 4.5).

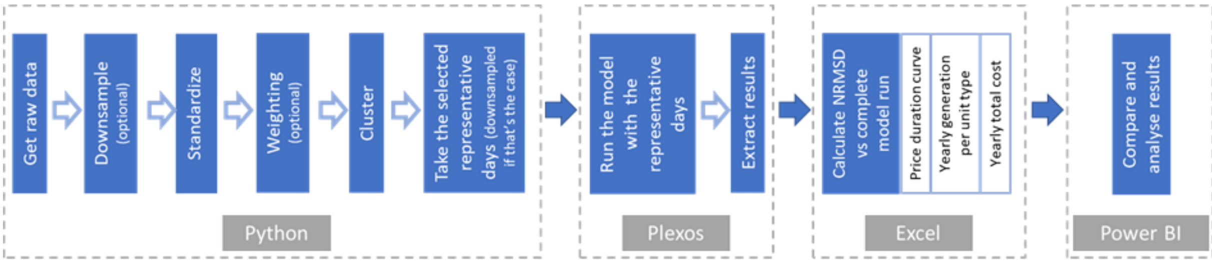

Figure 5 – Overview of process steps.

Figure 5 presents an overview of the process steps that had to be repeated for each different combination of weighting, downsampling, time seasonality (CET/UTC), and number of clusters. This was a rather time-consuming task, specially setting up the model with the different clustered input data, and extracting and comparing the results.

In the remainder of this section we outline each of these tasks in detail.

4.1. DATA COLLECTION AND ELECTRICITY MARKET MODEL DEVELOPMENT

The main objective of the modelling was to develop a representation of the Portuguese electricity system in order to apply already developed clustering techniques and also to explore new ways to aggregate intra-annual temporal variability of electricity demand data along with wind, solar and hydro availability. Thus, studying how different clustering algorithms and different partition sizes affect the accuracy of simulations on the Portuguese electricity system. The intention is to further understand which variables of the system can be accurately simulated taking only a fraction of the original simulation time and what methods can be used to improve this accuracy.

A simplified model of the Portuguese electricity system was developed using PLEXOS market simulation software. PLEXOS is a problem-solving engine, providing a single integrated hub for multiple systems and allowing modelling options from very simple to intricate and complex systems. PLEXOS translates the model runs into a series of linear-programming problems that then need a solver to be optimised. The solver package used for this study was the open-source GNU Linear Programming Kit[8].

[8] The model was created and ran on a laptop running with Windows 10 Pro 64bits with an Intel® Core™ i7-8550U CPU @ 1.80GHz processor and 16GB RAM.

The ultimate goal of this model was not to replicate the real system in the most complete and factual way, but rather to make it a working practical representation of the system, allowing the study of how reducing input data affects the model's outputs.

Portugal was modelled as a single region, assuming it to be as an isolated copper plate with no transmission system constraints, neglectable transmission losses and no cross-zonal trades.

Three groups of generation types were considered:

- combustion generators: fossil hard coal, fossil gas, and biomass[9];
- dispatchable renewables: hydro pumped storage, and hydro reservoir;
- intermittent renewables: solar, wind, and hydro run-of-river.

The generation capacity per unit type was replicated from European Network of Transmission System Operators for Electricity (ENTSO-E) Transparency Platform (TP) data (ENTSO-E, 2021). However, because the aim of this study was not to understand which generators are specifically dispatched, clustered unit commitment (CUC)[10] was used so that all different generators of the same type were grouped into an average one, having as many of these units as in the real system, and sticking to the real total maximum generation capacity described in TP. This meant that transmission system costs could also be disregarded for the purpose of this study. Since specific generation efficiency (heat rate) data was not readily and reliably available and fuel prices tend to be rather volatile, we considered standard and constant industry values for these variables.

For fossil fuelled generators, emission costs were implied in fuel costs instead of modelling emission costs separately, meaning that emission costs were considered constant along with fuel prices.

The model was developed for 5 years (2016-2020), taking in hourly data from ENTSO-E (2021) for Portuguese load[11], and solar, wind and hydro run-of-river generation.

According to ENTSO-E's TP data, hydro generation capacity represents over a third of Portugal's total capacity (see Figure 4). Hydro (with its subtypes of generation) has very particular characteristics, and misrepresenting it in a system where it plays such a large role can deeply affect the model's performance. In order to more accurately simulate hydro generation, hourly historical generation values from run-of-river units taken from TP were used as a proxy to natural inflow. For reservoir generators, since their generation can be controlled up to a certain point, natural inflow had to be calculated as hourly generation minus the hourly average of the difference between the week's final

[9] Biomass is considered a renewable energy source in the European Union assuming its fuel's origin is guaranteed to be sustainable. For the purpose of this study all biomass powerplants were considered equal in efficiency and fuel costs, rending irrelevant the fuel's source to the study's outcome. Furthermore, biomass was grouped into the combustion generators' group rather than dispatchable renewables' because of modelling similarity when disregarding costs of emissions, such as the case. See EU's biomass definition and sustainability criteria at: https://ec.europa.eu/energy/topics/renewable-energy/biomass_en

[10] CUC consists of grouping identical or similar powerplants to reduce the model's complexity by turning binary commitment variables of all plants within a group into a single integer variable. CUC has been shown to introduce very little error into power system models while reducing the computational cost, although grouping nonidentical generation units can increase the error (Meus, Poncelet, & Delarue, 2018).

[11] We used load and demand interchangeably since transmission losses were not modelled.

and initial stored energy[12] split between pumped hydro and hydro reservoir according to their relative generation capacity. For hydro pumped storage, natural inflow was calculated similarly to hydro reservoir's, plus a deduction of the pumped energy weighted by the pump efficiency, set at 75% as per industry standards. Hydro pumped storage and hydro reservoir's storage capacity were defined separately, and the maximum storage in historic data for the four years was split by the two types of generators according to their relative generation capacity.

Hydro pumped storage was the only modelled way to store generated energy for later consumption. This way the model should pump water upstream when prices are low enough to justify the 25% energy loss and generate when prices are higher.

In practical terms this means that the model has three intermittent generators with zero marginal cost: solar, wind and hydro run-of-river; a dispatchable zero marginal cost generator: hydro reservoir; a storage facility and dispatchable generator at either no marginal cost (from natural inflow) or at a marginal cost of the pumped energy divided by the pumping efficiency: hydro pumped storage; and three dispatchable combustion generators at three different marginal prices: fossil hard coal (22.5€/ Megawatt-hour) (MWh), biomass (48€/MWh), and fossil gas (64€/MWh) (see Table 1). These last three tend to represent what is known as baseload, an intermediate load step, and peak load prices, respectively.

The lack of data found on specifications of the Portuguese electricity market was a recurring problem, managed by making assumptions based on general benchmark values for the global markets.

The creation of a central provision of input data for modelling has been suggested by Wiese et al. (2018). In this publication the authors describe the benefits of open source centralised and harmonised energy modelling data. ENTSO-E's TP, which is the main data source of our study, was shown to be a great step towards this ideal situation, even though it still lacks detailed metadata, and format and source harmonisation. These were some of the constraints we faced to find and process the necessary data.

Prices and technical parameters were sourced as follows. The average operating heat rates of hard coal, natural gas and biomass power plants were derived based on TYNDP (2018), Open Power System Data (2017), and U.S. Energy Information Administration (2018). The price of natural gas was assumed to be approximately the non-household price declared for Portugal by Eurostat (2018). The price of coal was taken as €/tonne from Brito & Villalobos (2018) and converted to €/Gigajoule (GJ) at the rate of 1 tonne of coal equivalent equalling 29,3076 GJ. The price of biomass varies significantly between powerplants and depends on particular deals done with different suppliers, therefore the price of biomass was assumed based on knowledge from power plants operating with this source. The pumping efficiency of hydro power plants was established as an industry standard of 75% from Open Power System Data (2017). The minimum stable capacity was inferred from the calculations in TYNDP (2018) for hard coal and natural gas. Due to lack of conclusive data, the minimum stable capacity for hydro, biomass and wind power plants was considered to be 20% of maximum capacity, whereas solar was set to have no minimum stable capacity due to the technology's nature.

[12] TP's data on hydro storage is represented weekly and measured as potential energy rather than water volume. It is presented together for hydro pumped storage and hydro reservoir.

Table 1 - Combustion generation costs and properties

	Fuel price (€/GJ)	Heat rate (GJ/MWh)	Electricity cost (€/MWh)	Efficiency (%)	Minimum stable capacity (% of max)
Natural Gas	8	8	64	45%	35%
Hard Coal	2.5	9	22.5	40%	43%
Biomass	4	12	48	30%	20%

The possibility of wind turbines being curtailed was disregarded since it usually takes place due to energy imbalances or network constraints, neither being the object of this study. Furthermore, no feed-in tariffs or minimum running times were considered, and so no negative prices were expected to be observed.

Once all the data was collected to represent the Portuguese system and a PLEXOS model created with this data, we tested the validity of the model. To make sure the model was representative of the actual Portuguese market we ran it and compared prices and generation to the actual ones, fine-tuning fuel prices and storage capacity in order to make it as reliable as possible. The final model showed average hourly prices similar to the average Portuguese spot prices (OMIE, 2021) (see Figure 6), and coal, gas and biomass generation volumes in line with the actual ones (ENTSO-E, 2021) (see Figure 7).

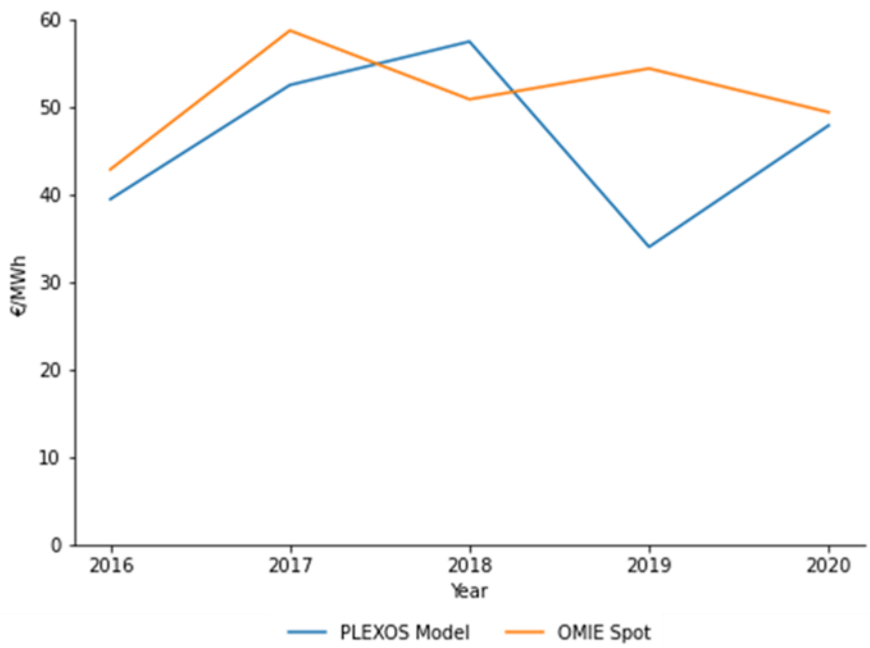

Figure 6 – Comparison between yearly average PLEXOS model price and OMIE Spot prices for Portugal (OMIE, 2021).

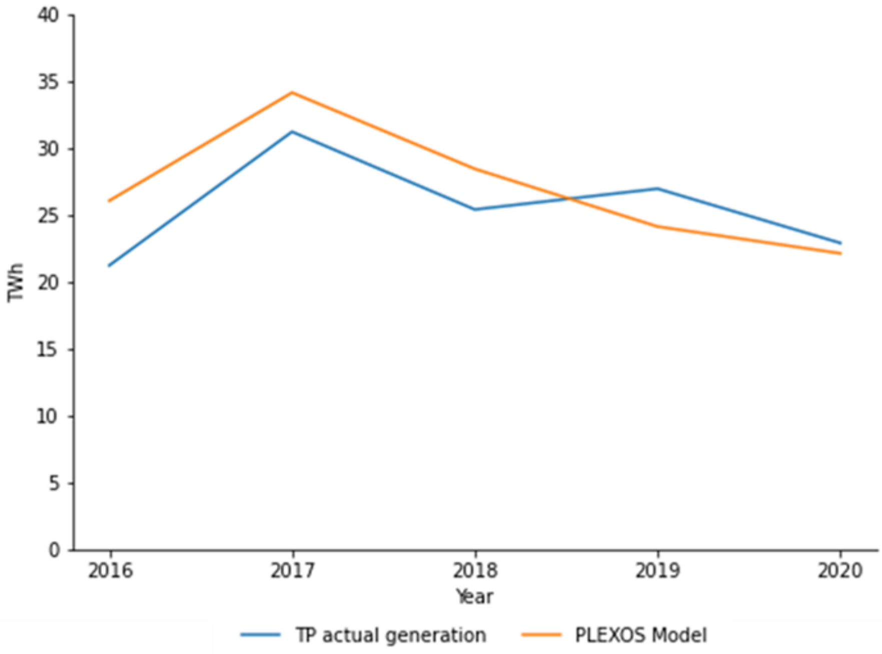

Figure 7 – Comparison between yearly total generation of biomass, gas, and coal from PLEXOS model and TP's actual generation.

4.2. DATA EXPLORATORY ANALYSIS

We then analysed the raw model input datasets to better understand their behaviour and the need for preprocessing. Each of the dimensions that comprise the dataset have their own specific yearly and daily profiles.

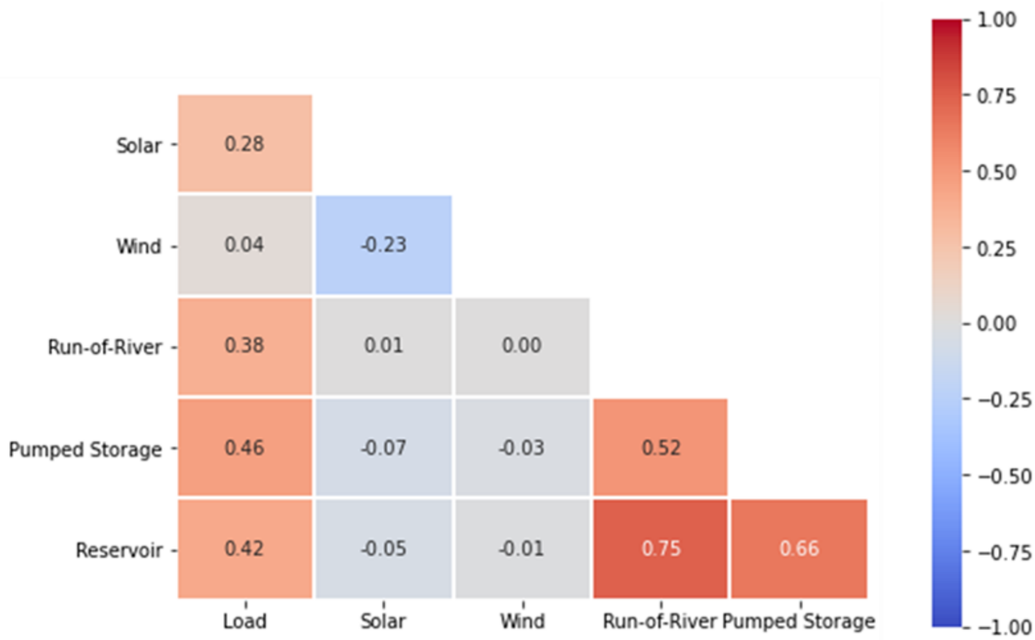

Figure 8 – Correlation matrix between the five-year hourly series of load, solar, wind, and the natural inflow of hydro run-of-river, pumped storage, and reservoir.

Figure 8 presents the correlation matrix between each of the six five-year hourly input series: load, solar, wind, and the natural inflow for hydro run-of-river, pumped storage, and reservoir. It shows load to have some positive correlation with solar generation and even more with the three natural inflows. Solar and wind generation have some negative correlation, and both have close to no correlation with the natural inflows. The natural inflows are highly positively correlated within themselves, which was to be expected by both their similar nature and way they were calculated (portrayed in detail in the previous section).

The lack of correlation between load, solar, and wind depicts a difficulty when trying to represent all the series in a limited number of days. In the remainder of this section we analyse the series' profiles, explaining their patterns and how they differ.

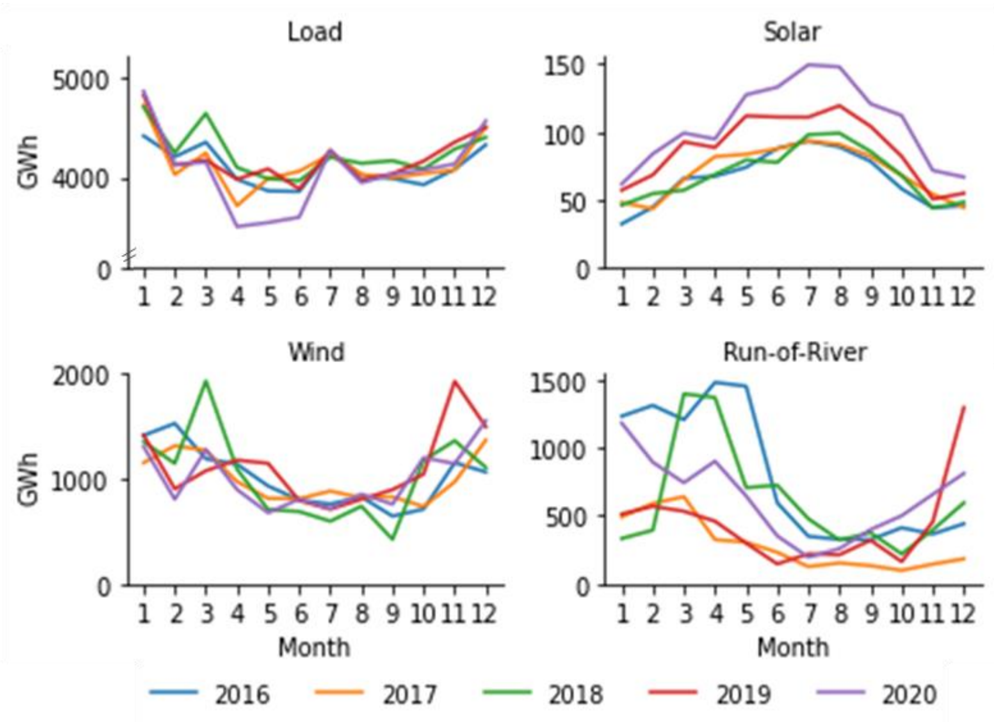

Figure 9 – Monthly total GWh of Load, Solar, Wind, and Hydro Run-of-River by year (2016-2020)

Figure 9 presents load, solar, wind, and hydro run-of-river yearly profiles as monthly total Gigawatt-hour (GWh) for each of the five years included in the study. From this frame of graphics we can analyse the overall profiles of each of the dimensions, and also the specific behaviour of each dimension in each year in Portugal.

Load has little seasonality, being slightly higher during winter months, despite the Portuguese winter not being considered a severe one compared to its summer, probably because of how poorly prepared Portuguese houses are for colder temperatures (Gouveia & Palma, 2021). The cold spell felt throughout Europe in late February to March 2018 (Copernicus, 2018) led to a peak demand month in March. Between April and June 2020, as the Covid pandemic hit, load faced a considerable plunge, having reached a 16-year low in April (Supiro, 2020).

Solar's average yearly profile shows a clear peak of generation during summer months and a plunge in winter months, with a smooth transition between them. From one year to the other the generation tends to be rather stable, and the increase in 2019 and again in 2020 are due to new generation capacity being built (as described in section 1) rather than any weather abnormality.

Wind presents an opposite yearly profile to solar, albeit not as pronounced, with the winters being windier than the summers. During the depicted years in particular, 2018 had a low wind summer, and there were wind generation peaks in March 2018 and November 2019.

Natural inflow, mostly encompassed by rain, is represented by the hydro run-of-river profile[13]. Similarly to wind generation, it tends it be higher during winter and also spring whilst lower in the summers. However, this is the dimension with the largest inter-annual variability, being largely influenced by periods of drenches and droughts. Water availability was high in 2016 up until May, whilst 2017 was a very dry year (Lusa, 2017). This drought ended in March 2018, which was the second rainiest March in Portugal in 87 years (Ferreira, 2018). This was followed by a dry 2019 up until November, having a very rainy December following suit to the start of 2020. Extreme weather events caused by climate change (Stott, 2016) are an ever more recurring problem, introducing difficulties when modelling power systems by increasing volatility and uncertainty.

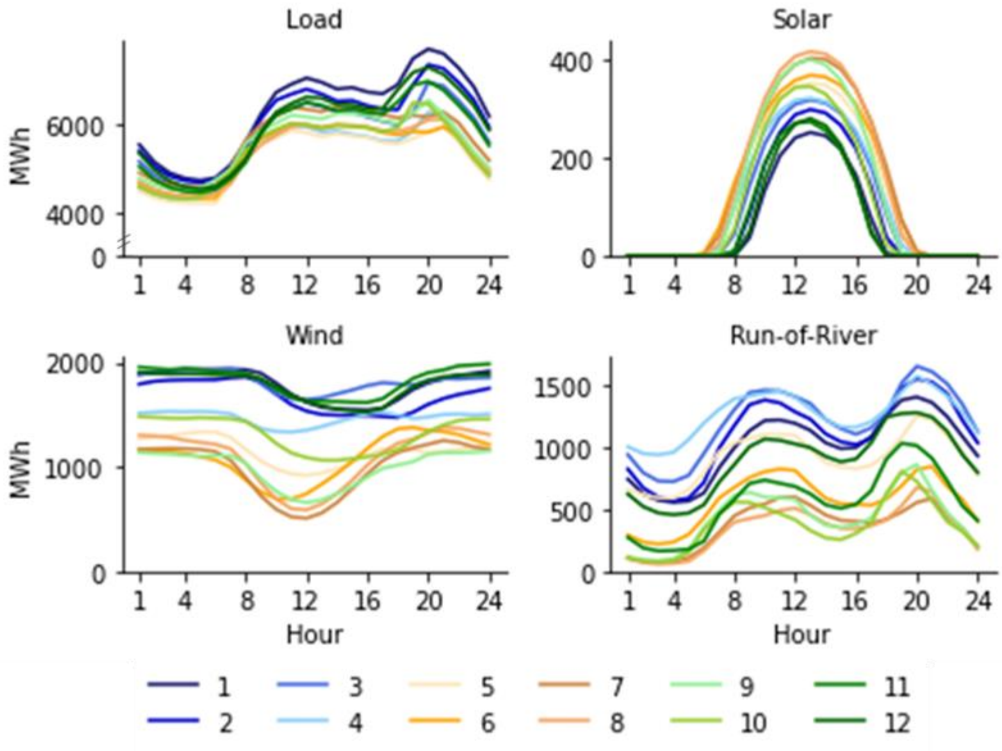

Figure 10 – Hourly average MWh daily profile in CET by month for load, solar, wind, and hydro run-of-river, average of 2016-2020.

Figure 10 presents the daily profiles of each of the four dimensions described in Figure 9, with the average MWh for each hour of the day, with a monthly detail (same months of different years averaged).

[13] Because of the way natural inflow for hydro pumped storage and hydro reservoirs was obtained as described in section 4.1, their profiles are very similar to hydro run-of-rivers, albeit with different magnitudes, as portrayed in Figure 8.

Load tends to peak around the start and end of the workday, with clear off-peak hours during the night. Winter's after-work peaks tend to be higher as heating systems are turned on, and summer curves tend to stagnate during working hours, as cooling systems during the hottest hours increase demand.

Solar electricity generation profile mimics the Sun's relative movement in the sky, being higher and out for longer during the summer, and lower and for a shorter period in the winter.

Wind is usually rather stable within each day, tending to slightly slow down around the warmest hours of the day. Once again it shows that warmer months are usually less windy.

Natural hydro inflow is usually higher during mornings and evenings, with the largest differentiation factor still being the time of the year, with much more hydro availability during the winter and spring.

From this we conclude that whereas representativeness of load and wind needs to focus more on intraday variance, for solar and hydro natural inflow the emphasis should be more on seasonality.

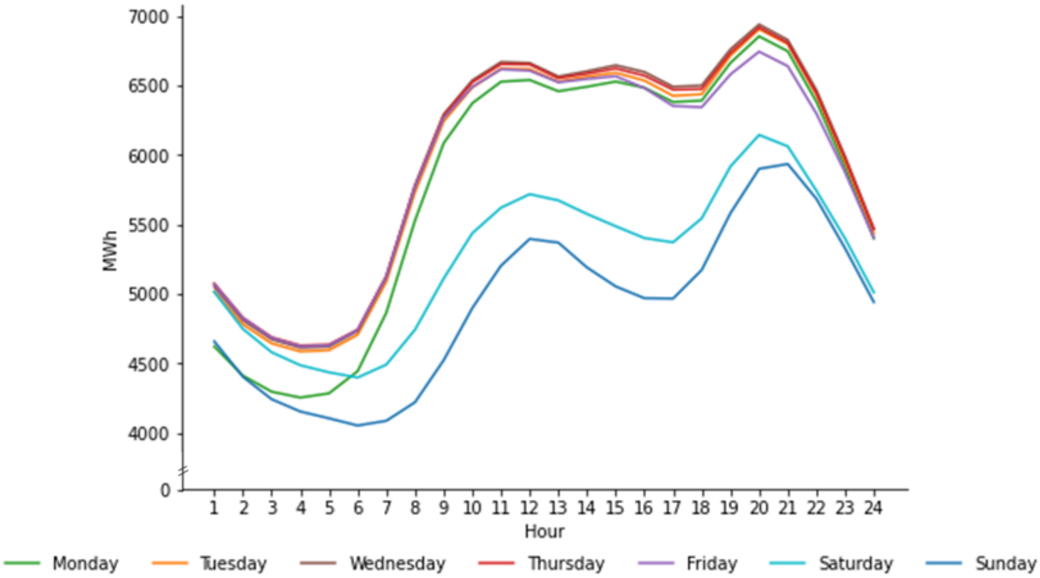

Figure 11 – Hourly average load MWh by weekday in CET, 2016-2020 Portugal.

As illustrated in Figure 11, load daily profile is also dependent on the weekday, being lower on weekends and peaking later in these mornings. This is not the case for other series, which means that focussing the clustering less on other dimensions besides just load should lead to less segregation between weekdays. This implies that to model demand, representative days need to account not just intraday and seasonal variances, but also intraweek ones.

4.3. Data Preprocessing

Having understood how our input datasets behave, we then have to process them for the clustering.

Separate standardization for each of the inputs (e.g., load, wind and solar) helps reducing the weight of the more stochastic wind series in the clustering algorithms, which presents much larger variations (i.e., point's distance) than load or solar, even though it is not necessarily more important to the model[14].

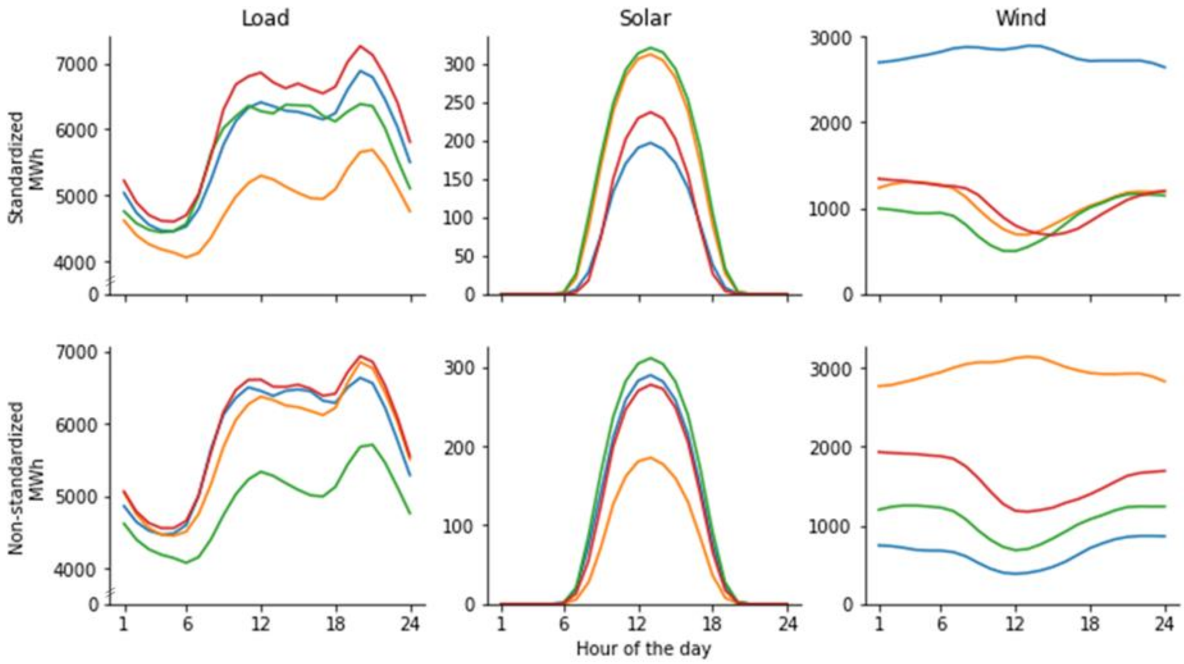

Figure 12 – Standardised and Non-Standardised Load, Solar, and Wind data clustered together using k-means with 4 clusters for 2016.

Figure 12 shows an example of how not standardizing the data can affect clustering results. It presents the results of the same k-means clustering into four clusters of 24-hour profiles of load, solar and wind, with the same input data preprocessed differently, providing a clear view on how the clustering effort is exogenously allocated when the data had been standardised by dimension (top row) compared to when the clustered data had not been standardised (bottom row). In Figure 12 line charts of the top row the load profiles were clustered into days with high, two medium (one with after working hours peak) and low demand, the solar profiles into days with more peak production and more hours of generation (summer-like profiles) and two days with less hours of production (winter-like profiles, one cloudier than the other), and the wind profiles were mostly separated into two groups (windy and non-windy days). On the other hand, in the bottom row load profiles were clustered only into high and low demand days, solar had much less clear winter/summer separation with mostly just a cloudier day, and wind was clustered into four visually different profiles. This shows that not standardizing the input data can lead to a larger focus on wind since it has more inter-day variability then the other dimensions.

As the original data comes from a reliable source and has no missing values[15], it is predicted that no outlier was produced by poor data quality. The outliers seen in Figure 13 are in fact an important aspect of the model input, especially the ones that represent hours of high net demand that will lead to blackouts if there is not enough installed capacity.

[14] This hypothesis was tested by weighting the different dimensions as described in Section 4.4

[15] Except for the daylight saving time changes in the CET datasets, where the two hours were averaged.

Since we intended to maintain the data's original shape and did not treat outliers, standardization was favoured over Max-Min Normalization (Bhandari, 2020).

The result was a data table with 144 columns which were the concatenation of the 24 hours in a day for each of the 6 dimensions, and as many rows as there were days for each of the years separately.

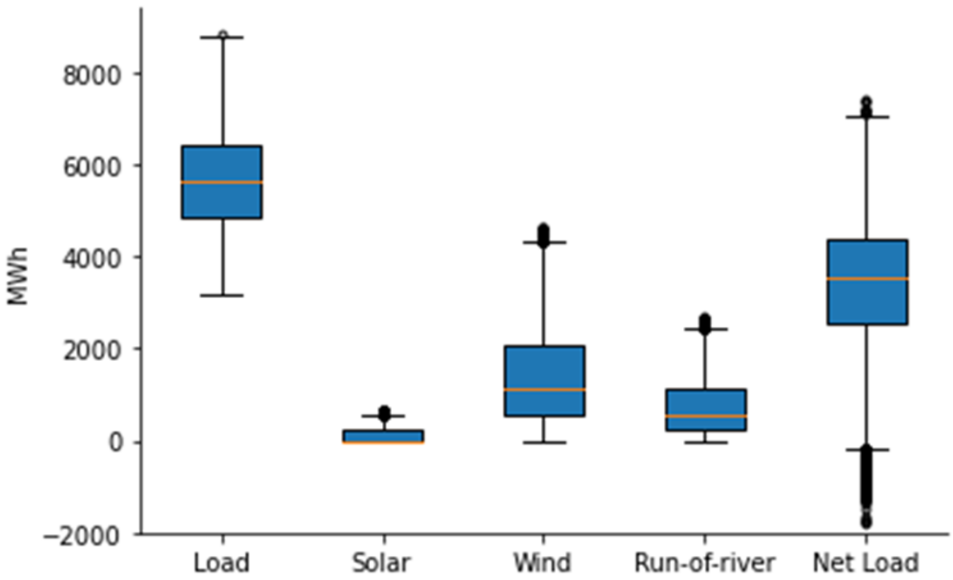

Figure 13 – Boxplots of load, solar, wind, and hydro run-of-river input datasets and of load net of solar, wind, and run-of-river.

4.4. Clustering

After the raw data for hourly load, solar generation, wind generation, hydro run-of-river generation, hydro pumped storage natural inflow, and hydro reservoir natural inflow were preprocessed, different clustering techniques were tried out and tested.

When clustering with k-means, the resulting centroid is an average value of input points, whereas with k-medoids we get an actual day as output. Because of this, k-means tends to overly smooth the data in a way that can make it unrealistic when examining intertemporal constraints in the modelled system. Thus, preventing the clustering from capturing the actual variability of the system.

Also bearing in mind that one of the goals was to cluster with multiple combinations of the six input data dimensions (load, solar, wind, hydro run-of-river, hydro pumped storage, and hydro reservoir) we opted to use k-medoids with Euclidean distance to cluster the data. This way ensuring we could use the medoids data of the same day for the dimensions that were not being used for the clustering. For example, when clustering only with load, solar and wind data, the data for the three hydro dimensions would be picked from each medoid's reference day. Others such as Heuberger, Staffell, Shah, & Dowell (2017) and Pineda & Morales (2018)have opted to cluster using k-means and after all iterations use the closest points to the centroids as medoids to avoid smoothing effects.

As it was shown in section 2, the most common procedure in similar studies is to either only use demand data or to also include solar and wind generation, with a tendency to include IRES as their contribution to the electricity systems increases. Since the largest renewable energy source in Portugal is hydro, we have decided to test clustering with load, solar, and wind (from here onwards referred to as clustering with 3 dimensions), to test with also intermittent hydro (adding hydro run-of-river to cluster with 4 dimensions), and finally to include the dispatchable hydro by adding the natural inflow of hydro pumped storage and hydro reservoir (clustering with 6 dimensions). All data was taken in hourly, except for the calculated natural inflow for hydro pumped storage and reservoir, which was weekly. In order to feed the model with these last two dimensions and still run it with hourly granularity, the weekly values were taken as an hourly average for all hours of each week.

Using the three hydro natural inflow data sets as separate dimensions could mean overly focusing on hydro data and not add much information to the clustering given that the three series are highly correlated (Figure 8). In order to understand which series contribute the most to the clustering without ruling any of them out, we have tested clustering with different weights for each dimension. Because the clustering used the Euclidean distance, the weighting was performed by multiplying the standardised values for each dimension by the square root of the weight. In practical terms, if the weights of all dimensions are the same there is no weighting, if the weight of a dimension is zero it is not being considered for the clustering, and if the weight of a dimension is 2 it is the same as including that dimension two times in the Euclidean distance (see demonstration in the annexes – section 9).

As the number of representative days can only be reduced to a certain amount while maintaining reasonable representativeness, we have also attempted to combine it with reducing intraday dimensionality, as suggested by Pfenninger (2017). This downsampling was achieved by averaging consecutive hours, easing the 24h different daily hours to 12, 8, 6, 4, 2, and 1 different one(s).[16]

Standardization and weighting were performed as follows:

$$\sqrt{w} \times \frac{x - \mu}{\sigma}$$

Where x is the input hourly value (already downsampled if that is the case), μ is the average of the set of x's hour and dimension, σ is the standard deviation of the set of x's hour and dimension, and w is the weight given to x's dimension. When downsampling was applied, x would be the average value of the intraday aggregated hours for x's dimension and day.

To perform and pipeline these tasks, a python script was developed, using the scikit-learn KMedoids package (scikit-learn, 2019) for clustering.

In an attempt to improve the efficiency of the clustering, parallelisation of the processing was introduced into the script using a multiprocessing python package. However, this package proved incompatible with the combination of the environment used to develop the scripts (Spyder for Windows) and modular programming, used to ensure code consistency and homogeneous maintenance of the scripts. This parallelisation strategy was dropped as it did not provide much added value to the study, since the recursive clustering was a minor part of the consumption of processing time and capacity, with the majority of it being used to take the clustered data and transform it into a format that the PLEXOS model could feed from and then extract and compare the outputs.

[16] Since the granularity of the natural inflow for hydro pumped storage and hydro reservoir is weekly (all daily values within a week are the same average ones due to raw data restrictions), there is no impact in changing the data granularity of these two dimensions.

4.5. ASSESSMENT OF CLUSTERING APPROACHES

In section 2 we have detailed how different comparison metrics have been used in previous studies, and the meaningfulness of their results. In this research we aimed not to understand how accurate our clustering was, but rather how representative in the model it was. This meant that instead of using metrics that compare the clustering, e.g., Euclidean distance between the original dataset and the cluster centre, more complex model output comparison metrics had to be devised.

The NRMSD of three comparison metrics was used to assess and compare the accuracy of each model run: hourly price duration curve, yearly generation by unit type, and yearly total generation cost. NRMSD for each of the comparison metrics was calculated as follows:

$$\frac{\sqrt{\frac{\sum_{i=1}^{n}(x_{ic} - x_{io})^2}{n}}}{\frac{\sum_{i=1}^{n} x_{io}}{n}}$$

Where x_{io} is the i -th value output of the original model run, x_{ic} is the i-th value output of the clustered model run, and n is the number of data points of the calculated comparison metric.

The hourly price duration curve is the price profile ordered descendant. It compares how well the model predicted the amount of time a certain market price was matched. This metric had 43848 data points[17] from each model run.

The yearly generation by unit type describes how much each of the 8 different generation unit types[18] produced in each year. This metric compares not only the total amounts generated but also how it was distributed between the different units. The metric had 40 data points[19] from each model run.

The third and last comparison metric - yearly generation cost - represents the total wholesale generation cost of the entire system for each of the five years (meaning 5 data points from each model run). We have also used this metric to understand how the NRMSD is distributed between the five years of the study, i.e., which years were more or less difficult to accurately cluster according to this metric.

We have decided to use these metrics because they represent key outputs for models to replicate. The above-mentioned metrics present an overview of price formation, unit commitment, and system costs, respectively.

We have also experimented clustering each dimension separately and then using the combinations of clusters of the various dimensions as the medoids, however this method proved to be both inefficient and unscalable.

Consecutively running the model with differently clustered data, extracting the results and comparing them was a rather laborious task. There were infinite dimension-weighting combinations that could be tested, along with downsampling and time seasonality. This had to be performed at a large-enough scale to have a decent variety of combinations to compare. The comparison would still be meaningful

[17] 365 days times 5 years (plus two days for the two non-leap years of 2016 and 2020) times 24h.
[18] Biomass, coal, gas, hydro pumped storage, hydro reservoir, hydro run-of-river, solar, and wind.
[19] 8 unit types times 5 years.

even with much less comparison points if we used a set of k clusters that showed some return and tended to not be very volatile. So, we decided to use only 6, 8, 10, 12, and 14 clusters to perform this task recursively and then run for 1-20, 25, 30, 40, 50, and 100 clusters for the most promising cases.

5. RESULTS AND DISCUSSION

In this section we review the results of the different clustering approaches for representative day selection. First we tested the behaviour of the three comparison metrics used in this study. Figure 14 provides a representative example of that, showing a tendency for the three metrics to correlate, with more similarities between the NRMSD of yearly generation by unit type and total yearly cost, while hourly price duration curve had overall larger errors. This points out that price is the most difficult and sensitive model output. Changing, for example, wind input, will directly affect wind and gas[20] production, but only indirectly have repercussions on the price if it changes the highest SRMC at that point in time.

There was some error volatility, with cases of lower k number of clusters having less error. This was to be expected when clustering with k-medoids, since it requires real data points for medoids. As the number of clusters increases, a larger part of the whole dataset is included, but not necessarily adding to the same days previously picked, which can affect sensitive outputs. Using the average of the NRMSD of the three comparison methods helped flatten this variance.

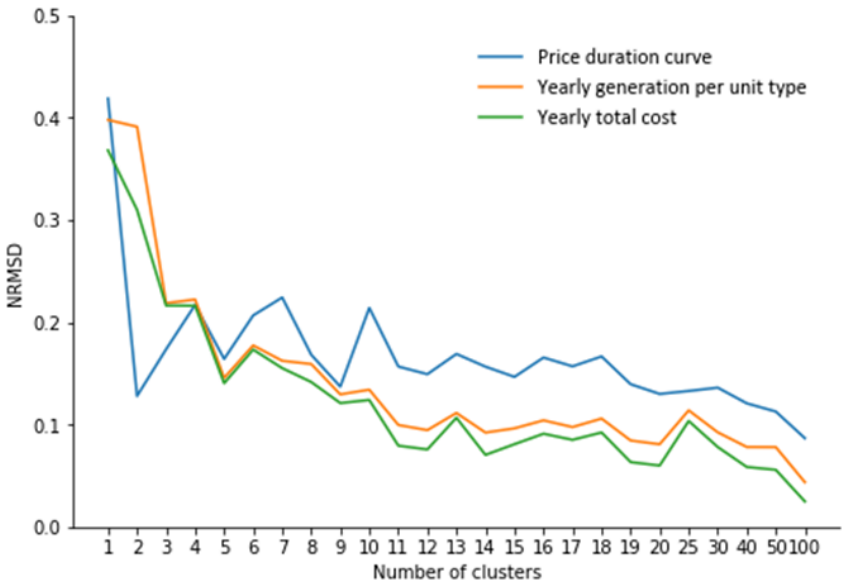

Figure 14 – NRMSD of price duration curve, generation by unit type, and total generation cost by clustering with load, solar, and wind data all with the same weight.

Figure 15 showcases an example of using Euclidean distance between original and clustered input datasets as accuracy metrics, opposed to the model output metrics we have used. In this example, results always get better as k number of clusters increases, since more of the total original dataset is used via medoids. Wind appears to be more difficult to cluster than load and solar, due it's the volatility mention in section 4.2. Even though the Euclidean distance for Solar seems to be very low, it does not mean it is sufficiently represented with only one day, emphasising the need to judge the clustering based on its effects on model outputs rather than its inputs.

[20] Assuming that net load is demanding all generators to produce at that point in time since gas was the last generator in the merit order.

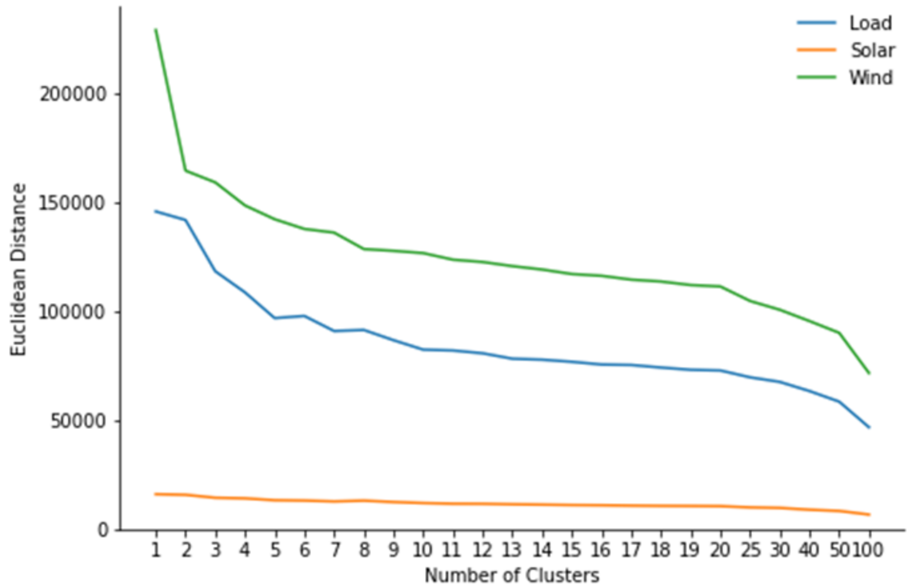

Figure 15 – Euclidean distance between original input data (of load, solar and wind) and the medoids from clustering with these dimensions all with the same weight, by number of clusters.

Figure 16 to Figure 18 showcase how clustering with data in CET versus UTC can lead to different model results, depending on the dimensions used to cluster. When clustering only with load, solar and wind data, time seasonality did appear to have an impact on the results (see Figure 16). However, when hydro was included in the clustering, this impact tended to fade, with the distinction becoming unclear (see Figure 17 and Figure 18).

Because of the data constraints on natural inflow for hydro pumped storage and reservoir (mentioned in section 4) these dimensions only have weekly variances. For this reason it came with no surprise that introducing these dimensions led to fading the differences between using CET and UTC data. However, this also happened when only introducing hydro run-of-river. This might have to do with natural river inflow not being linked to schedules but rather with natural events (e.g., temperature changes). This means that using input data in CET instead of UTC might not impact systems where the overwhelming presence of IRES reduces the relative importance of demand in the model.

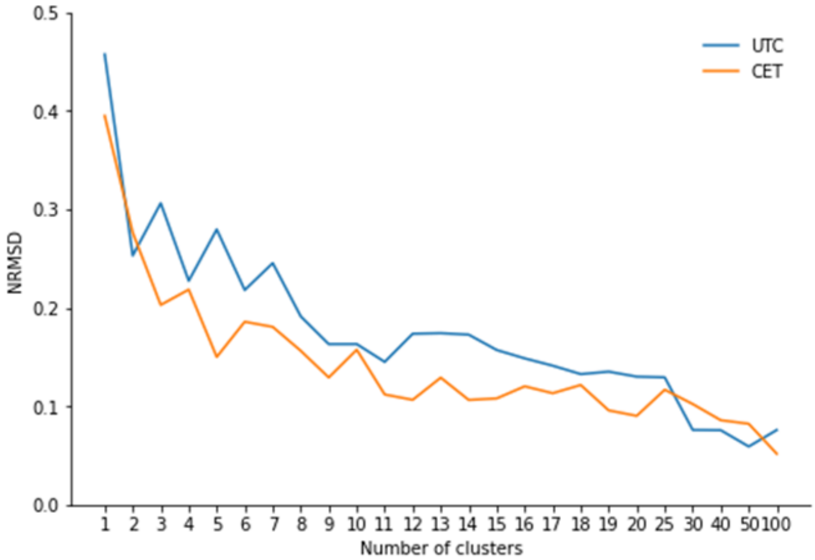

Figure 16 – Average NRMSD comparing clustering with CET vs UTC data, with 3 dimensions (load, solar, and wind) all with the same weight.

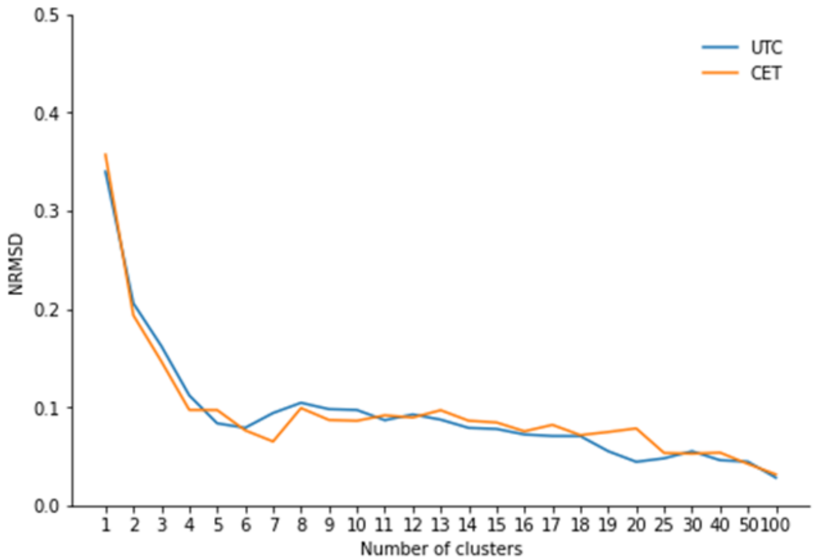

Figure 17 – Average NRMSD comparing clustering with CET vs UTC data, with 4 dimensions (load, solar, wind, and hydro run-of-river) all with the same weight.

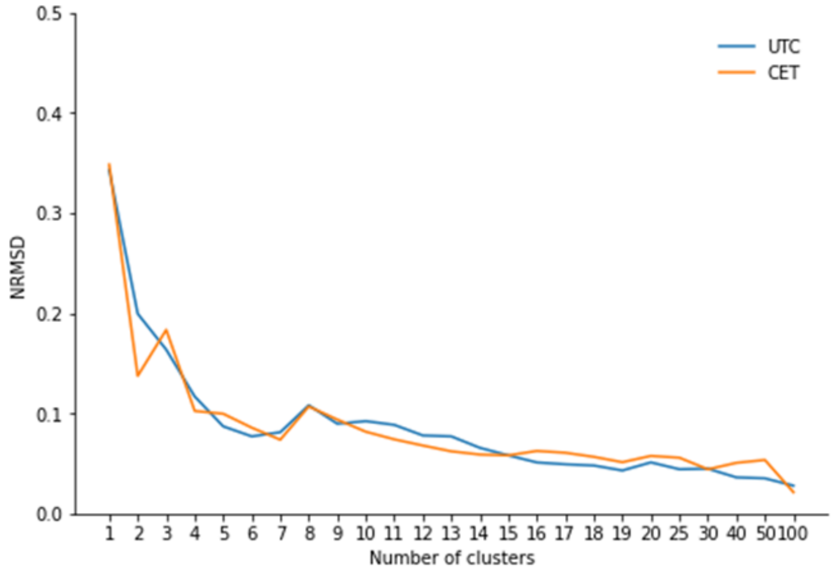

Figure 18 – Average NRMSD comparing clustering with CET vs UTC data, with 6 dimensions (load, solar, wind, hydro run-of-river, hydro pumped storage, and hydro reservoir) all with the same weight.

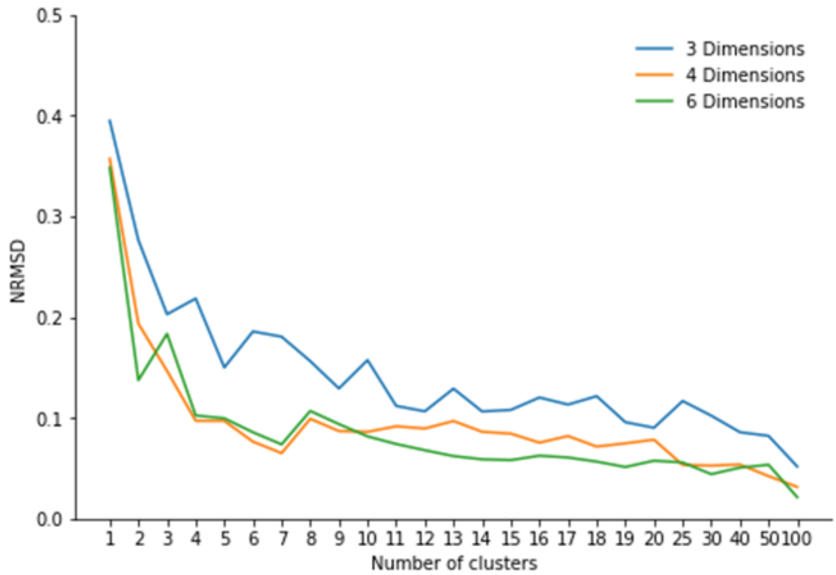

Figure 19 – Average NRMSD of price duration curve, generation by unit type, and total generation cost by k number of clusters (CET). Detail by dimensions used for clustering, all dimensions having the same weight.

Figure 19 shows the difference in average NRMSD of the three metrics using 3, 4, and 6 dimensions for clustering. Clustering with only load, solar, and wind (3 dimension) was consistently the worst option, even if not by a large margin. Clustering with 4 and 6 dimensions returned comparable results, although clustering with the 6 dimensions showed some separation with more than 10 clusters, converging again with over 20 clusters. This happens because with fewer k clusters the clustering chooses very similar days when considering 6 or 4 dimensions. However, as k increases, the clustering has more freedom to choose a larger variety of representative days, and because it has more detailed information with 6 dimensions than with 4, it tends to choose representative days more accurately. This shows that, if we can use a higher k number of clusters, it is useful to include multiple hydro series.

The average reduction on the average NRMSD of the three comparison metrics for k clusters between 1-20 was 32% from using 3 to 4 cluster dimension, and 11% from using 4 to 6 dimensions.

From these figures we can also infer that, even though the error tends to fall as k grows, the increase in accuracy becomes rather low with k larger than 12. Intuitively using the elbow method and also by analysing the decreasing rate-of-return, between 4 and 8 clusters would be considered the cut-off point since after that the return becomes residual. This number of days needed to efficiently represent the system is slightly lower than in most of similar studies (on different systems). To accurately model Great Britain's electricity system Green et al. (2014) used 6 to 10 representative days, while Pfenninger (2017) found the best trade-off between 10 to 15.

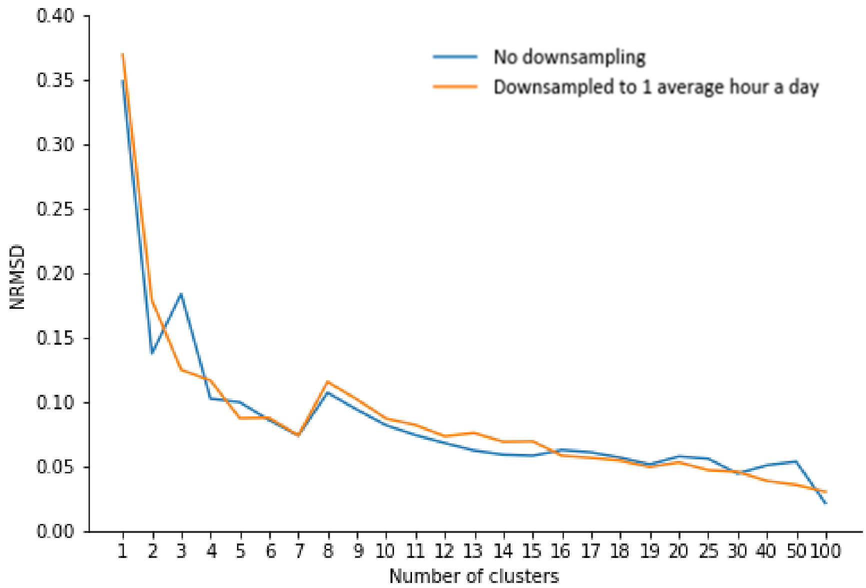

Figure 20 – Average NRMSD of price duration curve, generation by unit type, and total generation cost by number of clusters. Detail by downsampling, with all hours (CET) with original values versus all 24 hours in each day averaged out. Clustering with all 6 dimensions having the same weight.

Figure 20 portraits the effect of downsampling in the average NRMSD of the three comparison metrics. Reducing intraday granularity by averaging out consecutive hours had little to no effect on model results according to all three of the accuracy metrics used. Figure 21 depicts a similar comparison, but this time clustering with only 3 dimensions. While in this case downsampling introduced even more volatility to the results, the overall differences were a rather small and not conclusive decrease in accuracy. These results, although for a different system and using different metrics, oppose the conclusions from Pfenninger (2017) stating that downsampling consistently worsens results when considering high shares of IRES by smoothing peak demand.

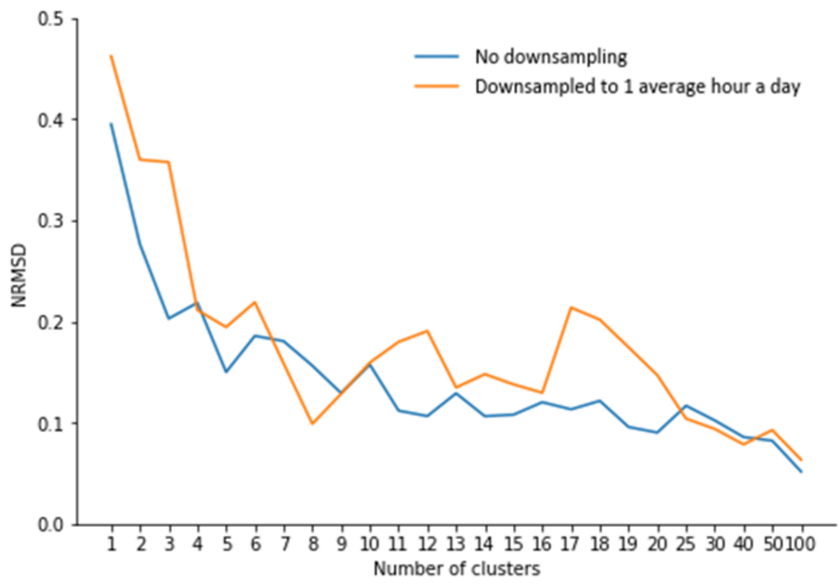

Figure 21 – Average NRMSD of price duration curve, generation by unit type, and total generation cost by number of clusters. Detail by downsampling, with all hours (CET) with original values versus all 24 hours in each day averaged out. Clustering with only load, solar and wind, the 3 dimensions having the same weight.

We have tested many different combinations of weighting for the various dimensions. In an attempt to pin down which dimensions positively (and negatively) affected the clustering accuracy, we doubled, halved, and increased by 50% the weight of each of the six dimensions individually. The results were not always consistent nor had a straight-forward interpretation. A more detailed analysis of the results (see all results in Table 4) led to the broad conclusion that results improved by increasing the weight of solar and wind data, and by decreasing the weight of hydro pumped storage and hydro reservoir (however the effect of reducing the weight of dispatchable hydro sources' natural inflow was not as conclusive).

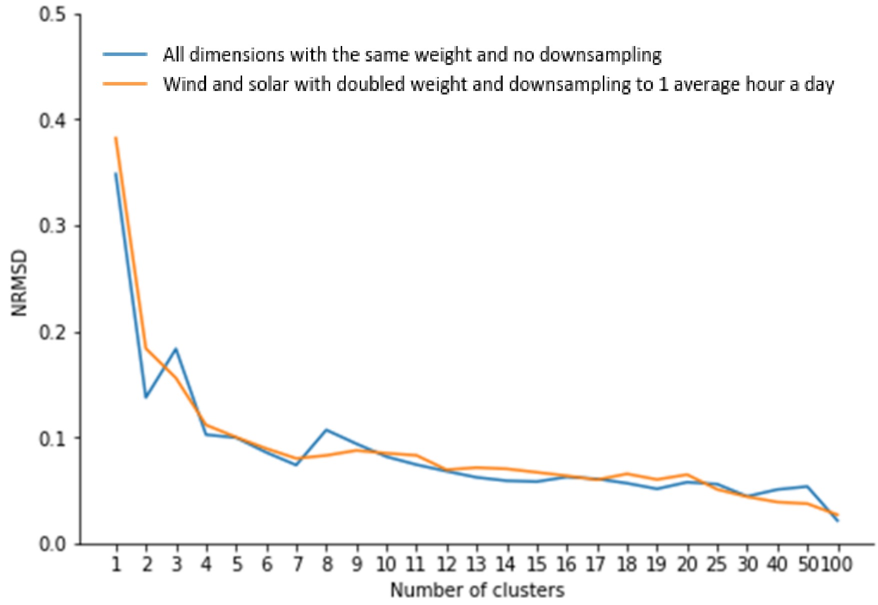

Figure 22 – Average NRMSD of price duration curve, generation by unit type, and total generation cost by number of clusters (CET). Comparison of no downsampling with all six dimensions with the same weight versus downsampling to one average value a day, and solar and wind with 50% more weight each.

Figure 22 presents the comparison of no downsampling and no weighting with the six dimensions (the unweighted approach) versus downsampling to the minimum one average value per day and emphasising solar and wind dimensions by 50% (the weighted and downsampled approach). The results of the later tend to be slightly better, while the first has much more volatility. This decrease in volatility is interesting, as it seems to imply that, without the weighting, unimportant series were changing the selected days significantly, which led to, by chance, improve results in some cases and in others worsen them.

Figure 23 presents a comparison of the Euclidean distances between the original load dataset and the load medoids from clustering with the same two techniques used in Figure 22, plus clustering using only 3 dimensions without weighting or downsampling. If we were to use this model input comparison instead of the model output metrics, the conclusions would be opposite to the ones previously described, once again showing that using model output comparison metrics is much more relevant to analyse.

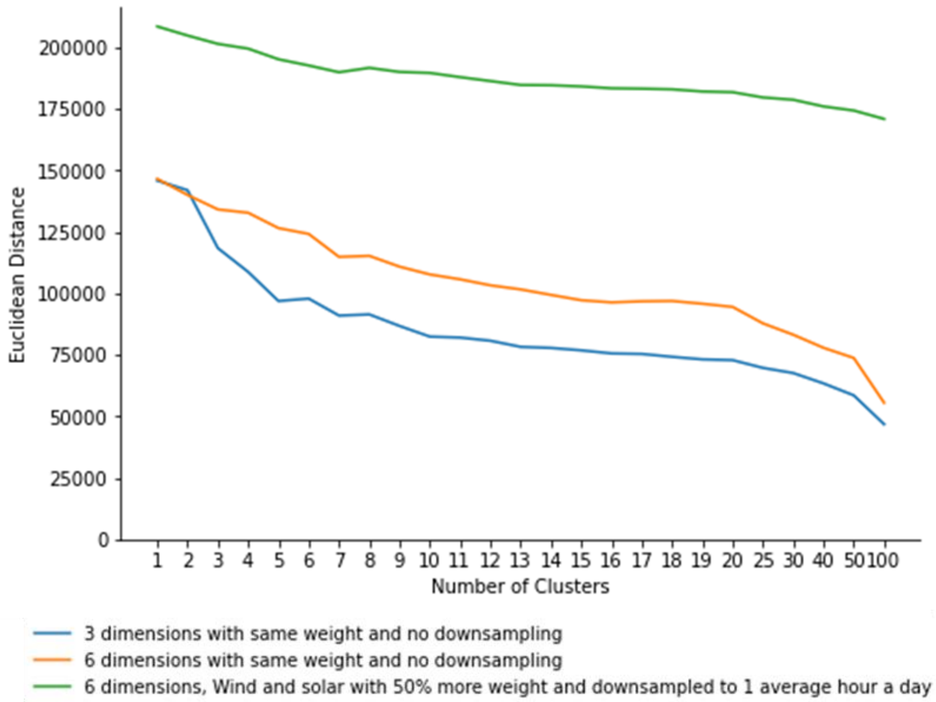

Figure 23 – Euclidean distance between original input data of load and the medoids from clustering with: three dimensions unweighted with no downsampling; six dimensions unweighted with no downsampling; and six dimensions downsampling to one average value a day.

Table 2 and Table 3 detail the distribution of the total cost NRMSD by each year for the unweighted approach (Table 2) and the weighted and downsampled approach (Table 3). From these tables we can infer that while the unweighted approach had much more difficulty modelling the costs for the year of 2020 for smaller k clusters and for 2017 for larger k (Table 2), the weighted and downsampled approach did not concentrate its error in any particular year (Table 3). A more detailed analysis into the yearly generation of each unit type lets us know that the unweighted method tended to underestimate hydro and wind generation, while not over-estimating solar enough to counter-balance, leading to higher combustion generation. The weighted and downsampled method estimated wind and solar generation more precisely, but still tended to underestimate hydro availability. As described in section 4.2, 2017 was a very dry year meaning that underestimating another significant RES source such as wind led to overestimating costs in the unweighted approach. 2020 was a rather odd year with a significant increase in solar and a couple of months of historically low demand due to lockdown, being hard to pin down the reason(s) for the poorer results.

Table 2 – Detailed model results of the distribution of the total cost NRMSD by each year for model runs with no downsampling and no weighting with the six dimensions (CET) with 1-20, 25, 30, 40, 50 k clusters.

Number of Clusters	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	27%	0%	25%	12%	36%
2	38%	2%	17%	4%	39%
3	17%	20%	11%	0%	51%
4	27%	28%	40%	4%	0%
5	21%	50%	25%	1%	3%
6	49%	0%	12%	25%	15%
7	14%	13%	16%	14%	43%
8	2%	6%	9%	30%	54%
9	3%	5%	9%	19%	63%
10	3%	12%	8%	12%	65%
11	5%	12%	16%	8%	58%
12	10%	22%	11%	1%	55%
13	7%	30%	3%	10%	50%
14	0%	12%	47%	1%	40%
15	0%	20%	16%	25%	39%
16	0%	43%	14%	9%	33%
17	0%	43%	17%	8%	32%
18	1%	59%	14%	1%	26%
19	1%	70%	22%	2%	5%
20	1%	87%	1%	9%	1%
25	5%	34%	58%	0%	2%
30	6%	15%	10%	8%	60%
40	0%	20%	49%	21%	10%
50	3%	41%	50%	3%	2%
100	0%	39%	53%	0%	7%

Table 3 – Detailed model results of the distribution of the total cost NRMSD by each year for model runs with downsampling to the minimum one average value per day and emphasising solar and wind dimensions by 50% (CET) with 1-20, 25, 30, 40, 50, and 100 k clusters.

Number of Clusters	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	59%	12%	17%	11%	1%
2	0%	6%	14%	58%	22%
3	37%	3%	0%	49%	11%
4	43%	6%	16%	8%	28%
5	23%	0%	9%	5%	63%
6	53%	12%	13%	0%	21%
7	74%	11%	15%	1%	0%
8	2%	56%	0%	38%	4%
9	2%	46%	8%	35%	9%
10	13%	47%	5%	30%	6%
11	0%	57%	6%	24%	13%
12	41%	4%	27%	17%	10%
13	9%	1%	60%	26%	4%
14	27%	12%	2%	60%	0%
15	24%	16%	7%	53%	1%
16	15%	44%	13%	11%	16%
17	4%	59%	15%	0%	22%
18	0%	48%	47%	1%	4%
19	32%	44%	13%	9%	2%
20	23%	16%	58%	3%	0%
25	15%	0%	36%	45%	4%
30	1%	3%	2%	34%	60%
40	8%	9%	8%	34%	42%
50	28%	19%	0%	27%	26%
100	3%	52%	37%	0%	8%

6. CONCLUSIONS

Modelling power systems allows for efficient data-driven decision-making. In order to test an immensity of hypothesis in said models, their complexity needs to be reduced, most commonly using representative days. We have studied ways to improve the accuracy of selecting representative days, in particular for the Portuguese electricity system. To achieve this goal we have modelled a representation of this system using PLEXOS, developed new clustering approaches and combined them with existing ones, selected model output comparison metrics, recursively ran the model with clustered inputs, and compared their results with the ones from running the model with the full dataset.

The results described in the section above enabled us to address the questions presented in section 3.

Firstly, we tried to understand how many representative days were needed to effectively model a year of the Portuguese power system (Q1). For the clustering techniques used, 4 to 8 representative days were enough to model the system with relatively low error. Further increasing the number of representative days had the benefit of reducing result volatility, while residually increasing accuracy.

Afterwards we studied the effect of using input data that ignored daylight saving time changes. Time seasonality impacted results when only load, solar, and wind were taken as clustering inputs, with data in CET outperforming data in UTC (Q2). When hydro was included in the clustering this difference faded and results were not conclusive. As more series that do not vary with society's routines were introduced, time seasonality's impact was minimized.

Because hydro has such a large impact in the Portuguese system, we investigated if it should be considered when clustering to find representative days (Q3). Introducing hydro run-over-river to the clustering input had a positive impact on results (32% average reduction on 1-20 k clusters for the average of the NRMSD for the three comparison metrics), as expected by the unusually large system dependency on hydro generation of over a third of total installed capacity. Further including natural inflow for hydro pumped storage and reservoir had a much more reduced impact (only -11% difference of the same average metrics comparing to including only 4 dimensions). However, this might have to do with the data limitations of the latter two dimensions.

We then applied the weighting technique we have developed to the clustering input data in order to study if all dimensions should be given the same weight, and how weighting optimises results (Q4). Many different combinations of weightings for the various input dimensions were tested, with only not very strong conclusions being drawn. Putting more weight on solar and wind input data (up to doubling it) tended to improve model results and reduce their volatility. The impact of different weightings is probably very case-specific, and even for the same system it might not continuously hold up. As IRES capacity continues to grow and change the relative importance of each of the unit types, their weights in the clustering should also be updated accordingly in order to avoid introducing some bias to the model.

Finally, we analysed the impact of intraday aggregations (downsampling) in the model's accuracy when combining it with other techniques (Q5). Reducing intraday granularity, in this case by averaging consecutive hours, presented very little to no negative impact on model output results. However, this complexity-reduction technique should be further studied in other models of the Portuguese system where all input data dimensions have full (hourly) granularity.

7. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

This study expanded the literature by investigating methods to create representative days for the Portuguese electricity system, combining different comparison metrics with multiple clustering, time seasonality, downsampling and weighting techniques. However, the research needed to make assumptions and overcome challenges that may have implications in the results. In addition, these findings suggest further areas for exploration.

Hydro generation is important for some countries like Portugal, however the available hydro data was limited because ENTSO-E's TP only publishes weekly hydro storage data, needed to calculate natural inflow for hydro pumped storage and hydro reservoir. We believe that having a hydro pumped storage and hydro reservoir dataset with less granularity (weekly) than the rest of the dataset (hourly) might have reduced the relevance of the conclusions taken when comparing clustering with all 6 dimensions. We thus reenforce the need for more complete, centralised, harmonised, and well-documented power system modelling data sources as presented by Wiese et al. (2018). As prosumers, V2G, and green energy demand become ever more frequent, the importance of including them in power system models also increases. However, this will only be possible with the publication of reliable, consistent, and well-documented data on these dimensions.

The three model accuracy measures used (NRMSD of hourly price duration curve, yearly generation per unit type, and yearly total costs) tended to be rather correlated between each other, indicating that all three metrics were broadly measuring model performance and not providing contradictory results. However, they also tended to have small NRMSD variation between different k number of clusters and clustering techniques, limiting the range of comparisons that could be made.

From this experience we conclude that more diversified comparison metrics could be tested for a more insightful comparison, eventually using metrics that take into account high and low peak events. Extreme events (at both ends of the spectrum) are the most difficult to model and can have great impacts on build decisions. For example, long periods of drought can mean that more dispatchable powerplants might need to be built in order to counterbalance IRES's volatility, as dispatchable hydro would not have enough stored energy for peak shaving.

Some of the aforementioned results' volatility could have been introduced by using k-medoids, so for future works we suggest comparing results with different clustering methods, namely by choosing a medoid only after all k-means iterations have ran, as performed by Pineda & Morales (2018). DTW could also be implemented with the mentioned clustering techniques. However, the datasets would need to be set out differently, as concatenating various dimensions the way we did it (see section 4.3) breaks the intertemporal sequence of a timeseries.

Differently weighting clustering input dimensions was shown to be able to improve modelling results. However, we were not able to study how case-specific these findings are, and what are the consequences of maintaining the weights of each dimension as a system evolves and changes its dynamics, namely when the penetration of IRES increases. In order to find optimal weights, a correlation between each dimension's characteristics and its weight could be established and compared in models with different capacity mixes. Ideally, investigating a formula for deciding the weight for each dimension based on their relative capacity, volatility, profile type, and other attributes.

8. BIBLIOGRAPHY

Balachandra, P., & Chandru, V. (1999). Modelling electricity demand with representative load curves. *Energy*, *24*(3), 219–230. https://doi.org/10.1016/S0360-5442(98)00096-6

Bhandari, A. (2020). *Feature Scaling for Machine Learning: Understanding the Difference Between Normalization vs. Standardization*. Retrieved from https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/

Booth, R. R. (1972). Power system simulation model based on probability analysis. *IEEE Transactions on Power Apparatus and Systems*, *PAS-91*(1), 62–69. https://doi.org/10.1109/TPAS.1972.293291

Brito, A., & Villalobos, L. (2018, June 15). Importações de carvão subiram para valor mais alto em 11 anos. *Público*. Retrieved from https://www.publico.pt/2018/06/15/economia/noticia/importacoes-de-carvao-subiram-para-valor-mais-alto-em-onze-anos-1834409

Clarke, L., Jiang, K., Akimoto, K., Babiker, M., Blanford, G., Fisher-Vanden, K., … Vuuren, D. P. van. (2014). Assessing Transformation Pathways. *Climate Change 2014 Mitigation of Climate Change*, *Chapter 6*(November), 413–510. https://doi.org/10.1017/cbo9781107415416.012

Copernicus. (2018). State-of-the-European-climate: February 2018. Retrieved June 20, 2021, from Copernicus State of the European Climate website: https://surfobs.climate.copernicus.eu/stateoftheclimate/february2018.php

Direção Geral de Energia e Geologia. (2021). *DGEG - Estatísticas Rápidas Renováveis n194*. Retrieved from https://www.dgeg.gov.pt/media/22eaof1k/dgeg-arr-2021-01.pdf

Elliston, B., MacGill, I., & Diesendorf, M. (2013). Least cost 100% renewable electricity scenarios in the Australian National Electricity Market. *Energy Policy*, *59*, 270–282. https://doi.org/10.1016/j.enpol.2013.03.038

Energias de Portugal. (2018). V2G Vehicle to Grid. Retrieved April 29, 2021, from Energias de Portugal website: https://www.edp.com/pt-pt/inovacao/v2g-vehicle-grid

ENTSO-E. (2021). ENTSO-E Transparency Platform. Retrieved September 16, 2020, from ENTSO-E Transparency Platform website: https://transparency.entsoe.eu/

ENTSO-E, & ENTSO-G. (2018). *TYNDP 2018 - Scenario Report*. Retrieved from https://tyndp.entsoe.eu/tyndp2018/scenario-report/

European Comission. (2015). EU Emissions Trading System. Retrieved April 21, 2021, from European Comission website: https://ec.europa.eu/clima/policies/ets_en

European Comission. (2019). Prosumers in the changing energy grid. Retrieved April 24, 2021, from European Comission CORDIS website: https://cordis.europa.eu/article/id/303126-prosumers-in-the-changing-energy-grid

Eurostat. (2018). Gas Prices. Retrieved September 16, 2020, from Eurostat website: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Gas_prices,_Second_semester_of_2016-2017_(EUR_per_kWh).png

Ferreira, A. B. (2018, April 2). Portugal já não está em seca, com o segundo março mais chuvoso em 87 anos. *Diário de Notícias*. Retrieved from https://www.dn.pt/portugal/portugal-ja-nao-esta-em-seca-com-o-segundo-marco-mais-chuvoso-em-87-anos-9229501.html

Figueira, J. (2018). *A History of Electricity in Portugal*. CTT.

Fortunato, A., Falcão, A., Domingues, A., Leite Garcia, A., Trindade, A., Grosa, C., … Santos, V. (2008). *A Regulação da Energia em Portugal 1997-2007* (Clube do Colecionador dos Correios, Ed.). ERSE.

Glowacki, M. (2016). Value of Lost Load. Retrieved April 21, 2021, from European Union Emissions Trading Scheme website: https://www.emissions-euets.com/internal-electricity-market-glossary/966-value-of-lost-load-voll

Götz, B., Blesl, M., Fahl, U., & Voß, A. (2012). *Theoretical background on the modeling of policy instruments in energy system models*.

Gouveia, J. P., & Palma, P. (2021). *Cold homes during the winter period and excess heat in homes during the summer in Portugal*. Retrieved from https://www.eppedia.eu/article/cold-homes-during-winter-period-and-excess-heat-homes-during-summer-portugal

Green, R., Staffell, I., & Vasilakos, N. (2014). Divide and Conquer? k-means clustering of demand data allows rapid and accurate simulations of the British electricity system. *IEEE Transactions on Engineering Management*, *61*(2), 251–260. https://doi.org/10.1109/TEM.2013.2284386

Hannah, L. (1979). *Electricity before Nationalisation*. Palgrave Macmillan, London. https://doi.org/https://doi.org/10.1007/978-1-349-03443-7

Härtel, P., Kristiansen, M., & Korpås, M. (2017). Assessing the impact of sampling and clustering techniques on offshore grid expansion planning. *Energy Procedia*, *137*, 152–161. https://doi.org/10.1016/j.egypro.2017.10.342

Hassan, S., Khosravi, A., Jaafar, J., & Raza, M. Q. (2014). Electricity load and price forecasting with influential factors in a deregulated power industry. *Proceedings of the 9th International Conference on System of Systems Engineering: The Socio-Technical Perspective, SoSE 2014*, 79–84. Glenelg, SA, Australia: IEEE. https://doi.org/10.1109/SYSOSE.2014.6892467

Heuberger, C. F., Staffell, I., Shah, N., & Dowell, N. Mac. (2017). A systems approach to quantifying the value of power generation and energy storage technologies in future electricity networks. *Computers and Chemical Engineering*, *107*, 247–256. https://doi.org/10.1016/j.compchemeng.2017.05.012

International Energy Agency. (2018). *CO2 emissions from fuel combustion*. Retrieved from https://www.iea.org/reports/co2-emissions-from-fuel-combustion-overview

International Energy Agency. (2019). *World Energy Investment*. Retrieved from https://www.iea.org/reports/world-energy-investment-2019

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

Krajačić, G., Duić, N., & Carvalho, M. da G. (2011). How to achieve a 100% RES electricity supply for Portugal? *Applied Energy*, *88*(2), 508–517. https://doi.org/10.1016/j.apenergy.2010.09.006

Kristiansen, M., Korpås, M., & Härtel, P. (2017). Sensitivity analysis of sampling and clustering techniques in expansion planning models. *Conference Proceedings - 2017 17th IEEE International*

Conference on Environment and Electrical Engineering and 2017 1st IEEE Industrial and Commercial Power Systems Europe, EEEIC / I and CPS Europe 2017, *609795*(609795). https://doi.org/10.1109/EEEIC.2017.7977727

Liu, Y., Sioshansi, R., & Conejo, A. J. (2018). Hierarchical Clustering to Find Representative Operating Periods for Capacity-Expansion Modeling. *IEEE Transactions on Power Systems*, *33*(3), 3029–3039. https://doi.org/10.1109/TPWRS.2017.2746379

Lusa. (2017, December 17). 2017: Seca fora do normal prejudica culturas e leva a medidas de poupança de água. *Diário de Notícias*. Retrieved from https://www.dn.pt/lusa/2017-seca-fora-do-normal-prejudica-culturas-e-leva-a-medidas-de-poupanca-de-agua-8991813.html

Matos, A. C. de, Mendes, F., Faria, F., & Cruz, L. (2004). *A Electricidade em Portugal - Dos primórdios à 2ª Guerra Mundial*. EDP Museu da Electricidade.

Mazumdar, M., & Chrzan, L. (1995). Monte Carlo stochastic simulation of electric power generation system production costs under time-dependent constraints. *Electric Power Systems Research*, *35*(2), 101–108. https://doi.org/10.1016/0378-7796(95)00994-9

Merrick, J. H. (2016). On representation of temporal variability in electricity capacity planning models. *Energy Economics*, *59*, 261–274. https://doi.org/10.1016/j.eneco.2016.08.001

Messner, S., & Schrattenholzer, L. (2000). MESSAGE-MACRO: Linking an energy supply model with a macroeconomic module and solving it iteratively. *Energy*, *25*(3), 267–282. https://doi.org/10.1016/S0360-5442(99)00063-8

Meus, J., Poncelet, K., & Delarue, E. (2018). Applicability of a Clustered Unit Commitment Model in Power System Modeling. *IEEE Transactions on Power Systems*, *33*(2), 2195–2204. https://doi.org/10.1109/TPWRS.2017.2736441

Morales Pedraza, J. (2019). Current Status and Perspective in the Use of Coal for Electricity Generation in the North America Region. In *Conventional Energy in North America*. https://doi.org/10.1016/b978-0-12-814889-1.00004-8

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., … Wilbanks, T. J. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, *463*(7282), 747–756. https://doi.org/10.1038/nature08823

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

OMIE. (2021). *Evolution of the electricity market - Annual report 2020*. Retrieved from https://www.omie.es/sites/default/files/2021-01/informe_anual_2020_en.pdf

Open Power System Data. (2017). Data Sources. Retrieved September 16, 2020, from Open Power System Data website: https://open-power-system-data.org/data-sources

Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, *44*(3), 678–693. https://doi.org/10.1016/j.patcog.2010.09.013

Pfenninger, S. (2017). Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability. *Applied Energy*, *197*, 1–13.

https://doi.org/10.1016/j.apenergy.2017.03.051

Pillai, J. R., & Bak-Jensen, B. (2010). Impacts of electric vehicle loads on power distribution systems. *2010 IEEE Vehicle Power and Propulsion Conference, VPPC 2010*. https://doi.org/10.1109/VPPC.2010.5729191

Pina, A., Silva, C., & Ferrão, P. (2011). Modeling hourly electricity dynamics for policy making in long-term scenarios. *Energy Policy*, *39*(9), 4692–4702. https://doi.org/10.1016/j.enpol.2011.06.062

Pineda, S., & Morales, J. M. (2018). Chronological time-period clustering for optimal capacity expansion planning with storage. *IEEE Transactions on Power Systems*, *33*(6), 7162–7170. https://doi.org/10.1109/TPWRS.2018.2842093

Poncelet, K. (2018). *Long-term energy-system optimization models*. KU Leuven.

Poncelet, K., Hoschle, H., Delarue, E., Virag, A., & Drhaeseleer, W. (2017). Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. *IEEE Transactions on Power Systems*, *32*(3), 1936–1948. https://doi.org/10.1109/TPWRS.2016.2596803

Prado, M. (2020, October 12). Energia solar: Instalação de novas centrais fotovoltaicas terá ano recorde. *Jornal Expresso*. Retrieved from https://expresso.pt/orcamento-estado/2020-10-12-Energia-solar-Instalacao-de-novas-centrais-fotovoltaicas-tera-ano-recorde

Šajn, N. (2016). Electricity "Prosumers." *European Parlamentary Research Service*, (November), 1–10. Retrieved from http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/593518/EPRS_BRI(2016)593518_EN.pdf

Sakoe, H., & Chiba, S. (1971). A dynamic programming approach to continuous speech recognition. *Proceedings of the Seventh International Congress on Acoustics*, *3*, 65–69. Budapest.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *26*(1), 43–49. https://doi.org/10.1109/TASSP.1978.1163055

scikit-learn. (2019). KMedoids. Retrieved June 19, 2021, from scikit-learn website: https://scikit-learn-extra.readthedocs.io/en/stable/generated/sklearn_extra.cluster.KMedoids.html

Scott, I. J. (2021). *Long-Term Uncertainty In Electricity Market Modelling, Policy, and Planning*. https://doi.org/10.31224/osf.io/mvqhz

Scott, I. J., Carvalho, P. M. S., Botterud, A., & Silva, C. A. (2019). Clustering representative days for power systems generation expansion planning: Capturing the effects of variable renewables and energy storage. *Applied Energy*, *253*(July), 113603. https://doi.org/10.1016/j.apenergy.2019.113603

SGS. (2021). Comunicação V2G. Retrieved April 29, 2021, from SGS website: https://www.sgs.pt/pt-pt/transportation/automotive/electrical-components/power-electronics/charging-stations/v2g-communication

Sisternes, F. J. de, & Webster, M. D. (2013). Optimal selection of sample weeks for approximating the net load in generation planning problems optimal selection of sample weeks for approximating the net load in generation planning problems. *Massachusetts Institute of Technology. Engineering Systems Division*, (January), 12. Retrieved from https://dspace.mit.edu/bitstream/handle/1721.1/102959/esd-wp-2013-03.pdf?sequence=1&isAllowed=y

Stott, P. (2016). How climate change affects extreme weather events. *Science*, *352*(6293), 1517–1518. https://doi.org/10.1126/science.aaf7271

Supiro, A. (2020, May 5). Consumo de eletricidade caiu 12% em abril para mínimo de 16 anos. *Observador*. Retrieved from https://observador.pt/2020/05/05/consumo-de-eletricidade-caiu-12-em-abril-para-minimo-de-16-anos/

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

U.S. Energy Information Administration. (2018). Average Operating Heat Rate for Selected Energy Sources. Retrieved September 20, 2020, from U.S. Energy Information Administration website: https://www.eia.gov/electricity/annual/html/epa_08_01.html

Vasconcelos, J. (2019). *A energia em Portugal*. Fundação Francisco Manuel dos Santos.

Wagner, L. (2013). Overview of Energy Storage Technologies. *Future Energy: Improved, Sustainable and Clean Options for Our Planet*, 613–631. https://doi.org/10.1016/B978-0-08-099424-6.00027-2

Wiese, F., Schlecht, I., Bunke, W. D., Gerbaulet, C., Hirth, L., Jahn, M., … Schill, W. P. (2018). Open Power System Data – Frictionless data for electricity system modelling. *Applied Energy*, *236*, 401–409. https://doi.org/10.1016/j.apenergy.2018.11.097

Yeganefar, A., Amin-Naseri, M. R., & Sheikh-El-Eslami, M. K. (2020). Improvement of representative days selection in power system planning by incorporating the extreme days of the net load to take account of the variability and intermittency of renewable resources. *Applied Energy*, *272*(March), 115224. https://doi.org/10.1016/j.apenergy.2020.115224

9. ANNEXES

The demonstration bellow demonstrates how weighting dimensions in the Euclidean distance is equivalent to multiplying the values of a dimension by the square root of the weight right before calculating the Euclidean distance. This example showcases that including a dimension twice in the Euclidean distance is equivalent to weighting the same dimension by 2 using the square root of the weight.

$$d(a_i, b_i) = \sqrt{(a_{i,x} - b_{i,x})^2 + (a_{i,y} - b_{i,y})^2 + (a_{i,y} - b_{i,y})^2} = \sqrt{(a_{i,x} - b_{i,x})^2 + 2 * (a_{i,y} - b_{i,y})^2}$$
$$= \sqrt{(a_{i,x} - b_{i,x})^2 + (\sqrt{2}(a_{i,y} - b_{i,y}))^2} = \sqrt{(a_{i,x} - b_{i,x})^2 + (\sqrt{2} * a_{i,y} - \sqrt{2} * b_{i,y}))^2}$$

Where i is the hour of the days a and b, for the dimensions x and y.

Table 4 – Detailed model results for all model runs with clustered data

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
0,5	1	1	1	1	1	24	CET	6	0,121	0,078	0,092	0,097	0%	9%	18%	2%	70%
0,5	1	1	1	1	1	24	CET	8	0,100	0,075	0,054	0,076	0%	28%	59%	6%	7%
0,5	1	1	1	1	1	24	CET	10	0,067	0,050	0,041	0,053	5%	46%	47%	1%	1%
0,5	1	1	1	1	1	24	CET	12	0,081	0,051	0,049	0,061	5%	31%	54%	2%	9%
0,5	1	1	1	1	1	24	CET	14	0,067	0,036	0,028	0,044	29%	0%	1%	39%	31%
0,5	1,5	1,5	1	1	1	24	CET	6	0,119	0,091	0,101	0,103	14%	32%	34%	16%	5%
0,5	1,5	1,5	1	1	1	24	CET	8	0,124	0,081	0,085	0,097	9%	8%	71%	1%	11%
0,5	1,5	1,5	1	1	1	24	CET	10	0,123	0,085	0,090	0,099	2%	1%	61%	21%	15%
0,5	1,5	1,5	1	1	1	24	CET	12	0,114	0,068	0,074	0,085	7%	6%	51%	31%	5%
0,5	1,5	1,5	1	1	1	24	CET	14	0,112	0,063	0,066	0,080	27%	1%	41%	20%	11%
1	0,5	1	1	1	1	24	CET	6	0,137	0,087	0,055	0,093	1%	0%	3%	0%	96%
1	0,5	1	1	1	1	24	CET	8	0,140	0,070	0,047	0,086	16%	0%	17%	5%	61%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	0,5	1	1	1	1	24	CET	10	0,106	0,068	0,053	0,075	26%	22%	10%	3%	38%
1	0,5	1	1	1	1	24	CET	12	0,100	0,058	0,035	0,064	36%	0%	8%	0%	55%
1	0,5	1	1	1	1	24	CET	14	0,096	0,051	0,045	0,064	25%	13%	7%	1%	54%
1	1	0,5	1	1	1	24	CET	6	0,147	0,122	0,097	0,122	8%	14%	0%	9%	68%
1	1	0,5	1	1	1	24	CET	8	0,140	0,098	0,086	0,108	7%	9%	4%	53%	27%
1	1	0,5	1	1	1	24	CET	10	0,132	0,087	0,081	0,100	1%	8%	16%	37%	38%
1	1	0,5	1	1	1	24	CET	12	0,124	0,074	0,074	0,091	5%	29%	8%	29%	29%
1	1	0,5	1	1	1	24	CET	14	0,116	0,075	0,071	0,087	6%	9%	27%	21%	38%
1	1	1	0	0	0	1	CET	1	0,445	0,483	0,458	0,462	43%	12%	3%	4%	39%
1	1	1	0	0	0	1	CET	2	0,445	0,315	0,319	0,360	49%	14%	0%	16%	21%
1	1	1	0	0	0	1	CET	3	0,445	0,321	0,307	0,358	48%	0%	0%	26%	26%
1	1	1	0	0	0	1	CET	4	0,202	0,224	0,208	0,211	15%	14%	25%	5%	41%
1	1	1	0	0	0	1	CET	5	0,234	0,213	0,137	0,195	1%	38%	38%	15%	7%
1	1	1	0	0	0	1	CET	6	0,254	0,212	0,191	0,219	39%	17%	34%	8%	1%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	0	0	0	1	CET	7	0,113	0,196	0,167	0,159	23%	9%	56%	2%	10%
1	1	1	0	0	0	1	CET	8	0,143	0,097	0,057	0,099	51%	5%	18%	6%	20%
1	1	1	0	0	0	1	CET	9	0,172	0,126	0,087	0,128	36%	0%	2%	8%	54%
1	1	1	0	0	0	1	CET	10	0,133	0,190	0,155	0,159	5%	3%	41%	2%	49%
1	1	1	0	0	0	1	CET	11	0,110	0,229	0,200	0,180	27%	2%	42%	1%	29%
1	1	1	0	0	0	1	CET	12	0,131	0,226	0,214	0,191	61%	0%	26%	10%	4%
1	1	1	0	0	0	1	CET	13	0,134	0,149	0,121	0,135	52%	0%	15%	26%	6%
1	1	1	0	0	0	1	CET	14	0,107	0,174	0,163	0,148	9%	1%	73%	13%	4%
1	1	1	0	0	0	1	CET	15	0,086	0,169	0,159	0,138	0%	1%	75%	21%	3%
1	1	1	0	0	0	1	CET	16	0,147	0,133	0,110	0,130	34%	0%	20%	27%	19%
1	1	1	0	0	0	1	CET	17	0,214	0,220	0,208	0,214	90%	0%	3%	2%	4%
1	1	1	0	0	0	1	CET	18	0,232	0,195	0,179	0,202	58%	0%	35%	2%	5%
1	1	1	0	0	0	1	CET	19	0,215	0,163	0,145	0,175	32%	7%	50%	5%	6%
1	1	1	0	0	0	1	CET	20	0,207	0,129	0,105	0,147	42%	4%	30%	10%	14%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	0	0	0	1	CET	25	0,138	0,096	0,078	0,104	39%	1%	29%	11%	20%
1	1	1	0	0	0	1	CET	30	0,098	0,103	0,081	0,094	0%	0%	2%	70%	27%
1	1	1	0	0	0	1	CET	40	0,088	0,085	0,062	0,079	15%	1%	22%	46%	16%
1	1	1	0	0	0	1	CET	50	0,105	0,094	0,079	0,093	15%	0%	28%	21%	36%
1	1	1	0	0	0	1	CET	100	0,091	0,067	0,032	0,063	24%	1%	4%	63%	8%
1	1	1	0	0	0	24	CET	1	0,419	0,398	0,368	0,395	66%	6%	16%	6%	6%
1	1	1	0	0	0	24	CET	2	0,128	0,391	0,310	0,276	34%	19%	30%	5%	12%
1	1	1	0	0	0	24	CET	3	0,174	0,219	0,216	0,203	20%	32%	40%	0%	9%
1	1	1	0	0	0	24	CET	4	0,217	0,222	0,216	0,219	56%	4%	33%	6%	1%
1	1	1	0	0	0	24	CET	5	0,164	0,146	0,140	0,150	61%	14%	0%	24%	0%
1	1	1	0	0	0	24	CET	6	0,207	0,177	0,173	0,186	51%	0%	44%	5%	0%
1	1	1	0	0	0	24	CET	7	0,224	0,162	0,155	0,181	45%	4%	31%	19%	1%
1	1	1	0	0	0	24	CET	8	0,168	0,159	0,142	0,156	18%	25%	29%	16%	12%
1	1	1	0	0	0	24	CET	9	0,137	0,130	0,121	0,129	0%	45%	21%	21%	14%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	0	0	0	24	CET	10	0,214	0,134	0,124	0,157	0%	30%	30%	30%	11%
1	1	1	0	0	0	24	CET	11	0,157	0,100	0,079	0,112	5%	4%	68%	7%	16%
1	1	1	0	0	0	24	CET	12	0,149	0,095	0,076	0,107	1%	2%	53%	42%	3%
1	1	1	0	0	0	24	CET	13	0,169	0,111	0,107	0,129	1%	1%	51%	38%	8%
1	1	1	0	0	0	24	CET	14	0,157	0,092	0,070	0,106	11%	1%	7%	73%	6%
1	1	1	0	0	0	24	CET	15	0,147	0,096	0,081	0,108	14%	16%	37%	18%	15%
1	1	1	0	0	0	24	CET	16	0,166	0,104	0,091	0,120	29%	0%	34%	15%	22%
1	1	1	0	0	0	24	CET	17	0,157	0,098	0,085	0,113	24%	0%	48%	2%	26%
1	1	1	0	0	0	24	CET	18	0,167	0,106	0,092	0,122	21%	0%	48%	5%	26%
1	1	1	0	0	0	24	CET	19	0,140	0,084	0,063	0,096	60%	2%	8%	12%	18%
1	1	1	0	0	0	24	CET	20	0,130	0,081	0,060	0,090	43%	0%	24%	13%	21%
1	1	1	0	0	0	24	CET	25	0,133	0,114	0,104	0,117	27%	31%	17%	12%	13%
1	1	1	0	0	0	24	CET	30	0,136	0,092	0,078	0,102	18%	32%	0%	19%	31%
1	1	1	0	0	0	24	CET	40	0,121	0,078	0,059	0,086	24%	34%	10%	0%	32%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	0	0	0	24	CET	50	0,113	0,078	0,056	0,082	59%	10%	1%	3%	27%
1	1	1	0	0	0	24	CET	100	0,087	0,044	0,025	0,052	20%	8%	3%	69%	0%
1	1	1	0	0	0	24	UTC	1	0,446	0,462	0,465	0,457	43%	18%	10%	3%	26%
1	1	1	0	0	0	24	UTC	2	0,271	0,259	0,229	0,253	44%	4%	27%	1%	24%
1	1	1	0	0	0	24	UTC	3	0,172	0,385	0,362	0,306	12%	16%	46%	7%	19%
1	1	1	0	0	0	24	UTC	4	0,231	0,236	0,216	0,227	49%	24%	11%	15%	0%
1	1	1	0	0	0	24	UTC	5	0,271	0,282	0,285	0,280	36%	20%	3%	15%	26%
1	1	1	0	0	0	24	UTC	6	0,236	0,212	0,207	0,218	53%	18%	2%	0%	28%
1	1	1	0	0	0	24	UTC	7	0,246	0,243	0,248	0,245	31%	12%	4%	8%	46%
1	1	1	0	0	0	24	UTC	8	0,235	0,174	0,165	0,191	16%	45%	18%	14%	7%
1	1	1	0	0	0	24	UTC	9	0,193	0,158	0,139	0,163	19%	15%	51%	4%	11%
1	1	1	0	0	0	24	UTC	10	0,159	0,169	0,161	0,163	37%	13%	46%	1%	3%
1	1	1	0	0	0	24	UTC	11	0,199	0,123	0,113	0,145	12%	13%	63%	5%	7%
1	1	1	0	0	0	24	UTC	12	0,158	0,184	0,179	0,174	34%	8%	40%	3%	15%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	0	0	0	24	UTC	13	0,204	0,161	0,158	0,174	6%	8%	59%	4%	24%
1	1	1	0	0	0	24	UTC	14	0,199	0,162	0,158	0,173	2%	9%	55%	8%	27%
1	1	1	0	0	0	24	UTC	15	0,192	0,140	0,139	0,157	3%	13%	57%	2%	26%
1	1	1	0	0	0	24	UTC	16	0,185	0,133	0,128	0,149	6%	13%	67%	2%	12%
1	1	1	0	0	0	24	UTC	17	0,172	0,130	0,122	0,141	0%	4%	82%	1%	13%
1	1	1	0	0	0	24	UTC	18	0,165	0,123	0,110	0,133	0%	6%	56%	14%	24%
1	1	1	0	0	0	24	UTC	19	0,163	0,129	0,114	0,135	2%	2%	60%	9%	28%
1	1	1	0	0	0	24	UTC	20	0,164	0,119	0,107	0,130	4%	1%	67%	1%	28%
1	1	1	0	0	0	24	UTC	25	0,154	0,118	0,116	0,129	83%	5%	7%	1%	5%
1	1	1	0	0	0	24	UTC	30	0,121	0,064	0,043	0,076	1%	14%	41%	10%	33%
1	1	1	0	0	0	24	UTC	40	0,101	0,071	0,055	0,076	12%	30%	7%	39%	12%
1	1	1	0	0	0	24	UTC	50	0,102	0,045	0,031	0,059	47%	22%	19%	12%	1%
1	1	1	0	0	0	24	UTC	100	0,098	0,073	0,057	0,076	20%	30%	3%	45%	3%
1	1	1	0,5	0,5	0,5	24	CET	6	0,118	0,089	0,064	0,090	37%	0%	33%	2%	28%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	0,5	0,5	0,5	24	CET	8	0,118	0,077	0,070	0,088	3%	8%	43%	2%	44%
1	1	1	0,5	0,5	0,5	24	CET	10	0,092	0,066	0,053	0,070	1%	0%	77%	5%	17%
1	1	1	0,5	0,5	0,5	24	CET	12	0,086	0,046	0,031	0,054	3%	1%	91%	4%	2%
1	1	1	0,5	0,5	0,5	24	CET	14	0,085	0,048	0,034	0,056	0%	17%	75%	8%	0%
1	1	1	0,5	1	1	24	CET	6	0,101	0,081	0,038	0,074	14%	1%	8%	73%	4%
1	1	1	0,5	1	1	24	CET	8	0,108	0,078	0,061	0,083	0%	19%	20%	32%	30%
1	1	1	0,5	1	1	24	CET	10	0,110	0,072	0,070	0,084	11%	18%	43%	25%	3%
1	1	1	0,5	1	1	24	CET	12	0,109	0,063	0,066	0,079	10%	20%	61%	8%	1%
1	1	1	0,5	1	1	24	CET	14	0,114	0,060	0,056	0,076	15%	12%	65%	4%	4%
1	1	1	1	0	0	1	CET	1	0,445	0,376	0,358	0,393	54%	11%	17%	13%	4%
1	1	1	1	0	0	1	CET	2	0,306	0,131	0,087	0,175	18%	5%	17%	10%	50%
1	1	1	1	0	0	1	CET	3	0,284	0,166	0,128	0,193	5%	13%	43%	34%	5%
1	1	1	1	0	0	1	CET	4	0,259	0,136	0,107	0,167	38%	39%	7%	4%	12%
1	1	1	1	0	0	1	CET	5	0,194	0,097	0,045	0,112	0%	55%	44%	1%	1%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	0	0	1	CET	6	0,154	0,114	0,061	0,110	1%	7%	1%	85%	6%
1	1	1	1	0	0	1	CET	7	0,137	0,079	0,057	0,091	29%	1%	16%	50%	4%
1	1	1	1	0	0	1	CET	8	0,178	0,088	0,042	0,103	31%	2%	24%	40%	3%
1	1	1	1	0	0	1	CET	9	0,149	0,086	0,038	0,091	44%	43%	2%	10%	0%
1	1	1	1	0	0	1	CET	10	0,142	0,113	0,068	0,107	30%	13%	20%	7%	29%
1	1	1	1	0	0	1	CET	11	0,117	0,110	0,089	0,106	1%	24%	41%	4%	30%
1	1	1	1	0	0	1	CET	12	0,107	0,060	0,025	0,064	27%	1%	3%	0%	69%
1	1	1	1	0	0	1	CET	13	0,105	0,057	0,035	0,065	31%	16%	16%	6%	31%
1	1	1	1	0	0	1	CET	14	0,094	0,060	0,036	0,063	9%	8%	68%	12%	3%
1	1	1	1	0	0	1	CET	15	0,100	0,068	0,052	0,073	15%	37%	8%	35%	5%
1	1	1	1	0	0	1	CET	16	0,117	0,065	0,048	0,076	25%	28%	15%	15%	17%
1	1	1	1	0	0	1	CET	17	0,090	0,057	0,040	0,062	0%	46%	44%	4%	5%
1	1	1	1	0	0	1	CET	18	0,097	0,049	0,030	0,059	0%	12%	36%	0%	52%
1	1	1	1	0	0	1	CET	19	0,105	0,054	0,026	0,062	10%	2%	39%	17%	33%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	0	0	1	CET	20	0,092	0,053	0,025	0,057	2%	13%	32%	26%	27%
1	1	1	1	0	0	1	CET	25	0,069	0,042	0,031	0,048	0%	3%	46%	3%	48%
1	1	1	1	0	0	1	CET	30	0,066	0,044	0,035	0,048	7%	3%	65%	2%	23%
1	1	1	1	0	0	1	CET	40	0,057	0,041	0,028	0,042	2%	5%	35%	1%	58%
1	1	1	1	0	0	1	CET	50	0,063	0,034	0,022	0,040	28%	15%	5%	7%	44%
1	1	1	1	0	0	1	CET	100	0,033	0,035	0,027	0,032	3%	5%	1%	57%	34%
1	1	1	1	0	0	24	CET	1	0,445	0,311	0,315	0,357	24%	4%	21%	10%	41%
1	1	1	1	0	0	24	CET	2	0,254	0,173	0,153	0,194	23%	8%	7%	6%	56%
1	1	1	1	0	0	24	CET	3	0,169	0,145	0,126	0,147	34%	5%	43%	6%	12%
1	1	1	1	0	0	24	CET	4	0,064	0,121	0,107	0,097	7%	0%	51%	9%	33%
1	1	1	1	0	0	24	CET	5	0,067	0,125	0,099	0,097	2%	18%	48%	7%	26%
1	1	1	1	0	0	24	CET	6	0,058	0,097	0,075	0,076	2%	51%	5%	8%	35%
1	1	1	1	0	0	24	CET	7	0,049	0,081	0,065	0,065	6%	26%	1%	4%	63%
1	1	1	1	0	0	24	CET	8	0,118	0,094	0,085	0,099	0%	15%	59%	25%	0%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	0	0	24	CET	9	0,125	0,076	0,060	0,087	8%	0%	77%	6%	8%
1	1	1	1	0	0	24	CET	10	0,091	0,090	0,077	0,086	13%	2%	45%	1%	39%
1	1	1	1	0	0	24	CET	11	0,089	0,099	0,087	0,092	11%	0%	31%	3%	55%
1	1	1	1	0	0	24	CET	12	0,102	0,089	0,077	0,089	20%	13%	17%	9%	41%
1	1	1	1	0	0	24	CET	13	0,107	0,096	0,088	0,097	20%	14%	23%	11%	32%
1	1	1	1	0	0	24	CET	14	0,103	0,085	0,071	0,086	9%	22%	31%	10%	27%
1	1	1	1	0	0	24	CET	15	0,101	0,083	0,070	0,084	9%	14%	33%	7%	37%
1	1	1	1	0	0	24	CET	16	0,085	0,078	0,063	0,075	3%	3%	33%	9%	51%
1	1	1	1	0	0	24	CET	17	0,084	0,084	0,077	0,082	5%	6%	39%	10%	40%
1	1	1	1	0	0	24	CET	18	0,084	0,069	0,062	0,072	1%	4%	55%	6%	35%
1	1	1	1	0	0	24	CET	19	0,076	0,077	0,072	0,075	0%	11%	40%	6%	44%
1	1	1	1	0	0	24	CET	20	0,078	0,082	0,076	0,078	2%	14%	25%	19%	40%
1	1	1	1	0	0	24	CET	25	0,071	0,047	0,041	0,053	2%	2%	50%	1%	46%
1	1	1	1	0	0	24	CET	30	0,071	0,047	0,039	0,053	6%	1%	51%	0%	42%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	0	0	24	CET	40	0,066	0,050	0,045	0,054	5%	15%	7%	7%	66%
1	1	1	1	0	0	24	CET	50	0,056	0,036	0,034	0,042	14%	31%	6%	9%	40%
1	1	1	1	0	0	24	CET	100	0,056	0,019	0,020	0,032	26%	1%	5%	38%	31%
1	1	1	1	0	0	24	UTC	1	0,439	0,291	0,289	0,340	25%	3%	25%	10%	38%
1	1	1	1	0	0	24	UTC	2	0,241	0,190	0,186	0,206	19%	27%	23%	2%	29%
1	1	1	1	0	0	24	UTC	3	0,169	0,167	0,151	0,162	1%	13%	54%	5%	27%
1	1	1	1	0	0	24	UTC	4	0,098	0,145	0,093	0,112	1%	9%	39%	40%	10%
1	1	1	1	0	0	24	UTC	5	0,069	0,112	0,069	0,084	21%	0%	26%	36%	17%
1	1	1	1	0	0	24	UTC	6	0,068	0,097	0,073	0,079	8%	37%	1%	6%	48%
1	1	1	1	0	0	24	UTC	7	0,093	0,108	0,080	0,094	11%	3%	60%	12%	15%
1	1	1	1	0	0	24	UTC	8	0,129	0,110	0,075	0,104	2%	5%	48%	44%	0%
1	1	1	1	0	0	24	UTC	9	0,122	0,099	0,073	0,098	2%	12%	35%	48%	4%
1	1	1	1	0	0	24	UTC	10	0,133	0,086	0,073	0,097	26%	0%	48%	16%	10%
1	1	1	1	0	0	24	UTC	11	0,114	0,080	0,066	0,087	27%	0%	47%	3%	23%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	0	0	24	UTC	12	0,126	0,079	0,073	0,093	25%	0%	69%	1%	5%
1	1	1	1	0	0	24	UTC	13	0,126	0,071	0,066	0,087	44%	0%	55%	1%	0%
1	1	1	1	0	0	24	UTC	14	0,120	0,066	0,051	0,079	18%	5%	55%	20%	0%
1	1	1	1	0	0	24	UTC	15	0,112	0,066	0,055	0,078	14%	8%	56%	20%	1%
1	1	1	1	0	0	24	UTC	16	0,097	0,062	0,058	0,072	6%	0%	78%	14%	2%
1	1	1	1	0	0	24	UTC	17	0,093	0,060	0,059	0,071	2%	3%	65%	20%	10%
1	1	1	1	0	0	24	UTC	18	0,090	0,061	0,061	0,071	6%	3%	55%	31%	5%
1	1	1	1	0	0	24	UTC	19	0,083	0,044	0,039	0,055	15%	20%	50%	12%	2%
1	1	1	1	0	0	24	UTC	20	0,054	0,040	0,039	0,044	4%	15%	38%	38%	6%
1	1	1	1	0	0	24	UTC	25	0,085	0,038	0,021	0,048	15%	51%	0%	25%	9%
1	1	1	1	0	0	24	UTC	30	0,092	0,042	0,031	0,055	10%	6%	72%	6%	6%
1	1	1	1	0	0	24	UTC	40	0,065	0,041	0,033	0,046	8%	4%	44%	13%	31%
1	1	1	1	0	0	24	UTC	50	0,068	0,038	0,028	0,045	0%	15%	64%	2%	19%
1	1	1	1	0	0	24	UTC	100	0,048	0,021	0,016	0,028	18%	19%	10%	3%	50%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	0,5	1	24	CET	6	0,109	0,078	0,037	0,075	7%	0%	7%	58%	29%
1	1	1	1	0,5	1	24	CET	8	0,101	0,071	0,050	0,074	3%	28%	23%	24%	22%
1	1	1	1	0,5	1	24	CET	10	0,107	0,058	0,058	0,075	1%	18%	48%	8%	25%
1	1	1	1	0,5	1	24	CET	12	0,087	0,045	0,040	0,057	9%	27%	42%	2%	20%
1	1	1	1	0,5	1	24	CET	14	0,080	0,042	0,030	0,051	22%	17%	28%	27%	6%
1	1	1	1	1	0,5	24	CET	6	0,109	0,078	0,037	0,075	7%	0%	5%	59%	29%
1	1	1	1	1	0,5	24	CET	8	0,102	0,071	0,051	0,075	3%	27%	22%	24%	24%
1	1	1	1	1	0,5	24	CET	10	0,107	0,058	0,058	0,075	1%	18%	48%	8%	25%
1	1	1	1	1	0,5	24	CET	12	0,087	0,045	0,040	0,057	9%	27%	42%	2%	20%
1	1	1	1	1	0,5	24	CET	14	0,080	0,039	0,032	0,050	19%	27%	24%	24%	5%
1	1	1	1	1	1	1	CET	1	0,445	0,331	0,332	0,369	67%	0%	19%	13%	1%
1	1	1	1	1	1	1	CET	2	0,251	0,154	0,130	0,179	1%	63%	12%	3%	22%
1	1	1	1	1	1	1	CET	3	0,203	0,105	0,066	0,125	9%	65%	19%	1%	6%
1	1	1	1	1	1	1	CET	4	0,129	0,111	0,110	0,116	2%	68%	28%	0%	1%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	1	1	1	CET	5	0,117	0,085	0,059	0,087	40%	21%	4%	2%	34%
1	1	1	1	1	1	1	CET	6	0,127	0,079	0,056	0,088	3%	87%	7%	1%	3%
1	1	1	1	1	1	1	CET	7	0,130	0,059	0,033	0,074	4%	75%	11%	3%	7%
1	1	1	1	1	1	1	CET	8	0,174	0,090	0,083	0,115	1%	10%	47%	14%	27%
1	1	1	1	1	1	1	CET	9	0,158	0,082	0,066	0,102	1%	5%	43%	23%	28%
1	1	1	1	1	1	1	CET	10	0,146	0,067	0,048	0,087	3%	5%	32%	7%	53%
1	1	1	1	1	1	1	CET	11	0,143	0,061	0,042	0,082	29%	5%	15%	9%	42%
1	1	1	1	1	1	1	CET	12	0,133	0,059	0,027	0,073	68%	0%	7%	3%	22%
1	1	1	1	1	1	1	CET	13	0,130	0,067	0,030	0,076	81%	13%	0%	2%	5%
1	1	1	1	1	1	1	CET	14	0,117	0,059	0,032	0,069	59%	23%	6%	12%	0%
1	1	1	1	1	1	1	CET	15	0,108	0,062	0,038	0,069	46%	32%	2%	8%	12%
1	1	1	1	1	1	1	CET	16	0,093	0,052	0,030	0,058	38%	10%	7%	1%	44%
1	1	1	1	1	1	1	CET	17	0,089	0,052	0,029	0,056	33%	6%	20%	12%	29%
1	1	1	1	1	1	1	CET	18	0,083	0,052	0,028	0,054	19%	8%	21%	13%	40%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	1	1	1	CET	19	0,084	0,046	0,018	0,049	16%	12%	35%	31%	6%
1	1	1	1	1	1	1	CET	20	0,091	0,045	0,023	0,053	5%	9%	4%	81%	1%
1	1	1	1	1	1	1	CET	25	0,080	0,041	0,020	0,047	32%	56%	4%	5%	3%
1	1	1	1	1	1	1	CET	30	0,078	0,037	0,022	0,046	45%	45%	0%	2%	8%
1	1	1	1	1	1	1	CET	40	0,072	0,030	0,013	0,039	2%	79%	1%	6%	13%
1	1	1	1	1	1	1	CET	50	0,060	0,031	0,015	0,035	6%	0%	47%	41%	7%
1	1	1	1	1	1	1	CET	100	0,059	0,022	0,010	0,030	0%	29%	17%	25%	29%
1	1	1	1	1	1	2	CET	6	0,156	0,088	0,088	0,111	8%	33%	9%	7%	42%
1	1	1	1	1	1	2	CET	8	0,155	0,066	0,061	0,094	4%	24%	8%	54%	11%
1	1	1	1	1	1	2	CET	10	0,115	0,050	0,024	0,063	38%	8%	1%	31%	21%
1	1	1	1	1	1	2	CET	12	0,114	0,053	0,035	0,067	43%	3%	13%	37%	4%
1	1	1	1	1	1	2	CET	14	0,111	0,052	0,040	0,068	24%	10%	29%	37%	0%
1	1	1	1	1	1	3	CET	6	0,112	0,081	0,066	0,086	27%	19%	6%	3%	44%
1	1	1	1	1	1	3	CET	8	0,101	0,061	0,038	0,067	19%	1%	31%	16%	33%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	1	1	3	CET	10	0,120	0,067	0,064	0,084	28%	0%	40%	29%	3%
1	1	1	1	1	1	3	CET	12	0,103	0,055	0,050	0,069	43%	5%	27%	22%	3%
1	1	1	1	1	1	3	CET	14	0,108	0,056	0,057	0,074	24%	26%	23%	22%	5%
1	1	1	1	1	1	4	CET	6	0,131	0,073	0,027	0,077	19%	37%	16%	22%	6%
1	1	1	1	1	1	4	CET	8	0,123	0,102	0,083	0,103	21%	20%	4%	33%	22%
1	1	1	1	1	1	4	CET	10	0,131	0,092	0,076	0,100	50%	7%	2%	25%	16%
1	1	1	1	1	1	4	CET	12	0,100	0,055	0,045	0,066	35%	11%	21%	32%	1%
1	1	1	1	1	1	4	CET	14	0,115	0,047	0,032	0,065	21%	6%	14%	45%	13%
1	1	1	1	1	1	6	CET	6	0,137	0,110	0,086	0,111	6%	27%	2%	12%	52%
1	1	1	1	1	1	6	CET	8	0,089	0,063	0,044	0,065	4%	19%	1%	53%	22%
1	1	1	1	1	1	6	CET	10	0,116	0,054	0,040	0,070	25%	8%	0%	25%	41%
1	1	1	1	1	1	6	CET	12	0,119	0,076	0,066	0,087	18%	48%	11%	1%	23%
1	1	1	1	1	1	6	CET	14	0,104	0,056	0,037	0,066	33%	41%	3%	3%	21%
1	1	1	1	1	1	8	CET	6	0,138	0,088	0,060	0,095	17%	3%	6%	26%	47%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	1	1	8	CET	8	0,109	0,062	0,043	0,071	0%	0%	7%	75%	18%
1	1	1	1	1	1	8	CET	10	0,116	0,063	0,056	0,078	1%	43%	1%	11%	43%
1	1	1	1	1	1	8	CET	12	0,096	0,056	0,045	0,066	0%	77%	6%	9%	7%
1	1	1	1	1	1	8	CET	14	0,092	0,041	0,025	0,053	0%	58%	3%	2%	37%
1	1	1	1	1	1	12	CET	6	0,122	0,074	0,050	0,082	69%	0%	18%	12%	0%
1	1	1	1	1	1	12	CET	8	0,104	0,066	0,058	0,076	2%	6%	1%	37%	53%
1	1	1	1	1	1	12	CET	10	0,115	0,071	0,062	0,083	13%	1%	3%	50%	34%
1	1	1	1	1	1	12	CET	12	0,108	0,060	0,058	0,076	15%	3%	26%	31%	26%
1	1	1	1	1	1	12	CET	14	0,115	0,067	0,061	0,081	9%	22%	31%	0%	38%
1	1	1	1	1	1	24	CET	1	0,445	0,307	0,293	0,348	27%	0%	25%	12%	36%
1	1	1	1	1	1	24	CET	2	0,164	0,137	0,111	0,137	38%	2%	17%	4%	39%
1	1	1	1	1	1	24	CET	3	0,233	0,167	0,150	0,183	17%	20%	11%	0%	51%
1	1	1	1	1	1	24	CET	4	0,149	0,091	0,068	0,102	27%	28%	40%	4%	0%
1	1	1	1	1	1	24	CET	5	0,107	0,109	0,083	0,100	21%	50%	25%	1%	3%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	1	1	24	CET	6	0,119	0,078	0,060	0,086	49%	0%	12%	25%	15%
1	1	1	1	1	1	24	CET	7	0,118	0,075	0,029	0,074	14%	13%	16%	14%	43%
1	1	1	1	1	1	24	CET	8	0,143	0,090	0,088	0,107	2%	6%	9%	30%	54%
1	1	1	1	1	1	24	CET	9	0,140	0,073	0,068	0,094	3%	5%	9%	19%	63%
1	1	1	1	1	1	24	CET	10	0,120	0,065	0,061	0,082	3%	12%	8%	12%	65%
1	1	1	1	1	1	24	CET	11	0,112	0,059	0,051	0,074	5%	12%	16%	8%	58%
1	1	1	1	1	1	24	CET	12	0,100	0,054	0,050	0,068	10%	22%	11%	1%	55%
1	1	1	1	1	1	24	CET	13	0,088	0,053	0,046	0,062	7%	30%	3%	10%	50%
1	1	1	1	1	1	24	CET	14	0,088	0,047	0,042	0,059	0%	12%	47%	1%	40%
1	1	1	1	1	1	24	CET	15	0,087	0,047	0,041	0,058	0%	20%	16%	25%	39%
1	1	1	1	1	1	24	CET	16	0,093	0,051	0,043	0,063	0%	43%	14%	9%	33%
1	1	1	1	1	1	24	CET	17	0,096	0,048	0,038	0,061	0%	43%	17%	8%	32%
1	1	1	1	1	1	24	CET	18	0,091	0,045	0,034	0,057	1%	59%	14%	1%	26%
1	1	1	1	1	1	24	CET	19	0,088	0,041	0,025	0,051	1%	70%	22%	2%	5%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	1	1	24	CET	20	0,089	0,050	0,034	0,058	1%	87%	1%	9%	1%
1	1	1	1	1	1	24	CET	25	0,089	0,041	0,038	0,056	5%	34%	58%	0%	2%
1	1	1	1	1	1	24	CET	30	0,073	0,035	0,025	0,044	6%	15%	10%	8%	60%
1	1	1	1	1	1	24	CET	40	0,077	0,038	0,037	0,051	0%	20%	49%	21%	10%
1	1	1	1	1	1	24	CET	50	0,075	0,044	0,041	0,054	3%	41%	50%	3%	2%
1	1	1	1	1	1	24	CET	100	0,039	0,014	0,011	0,021	0%	39%	53%	0%	7%
1	1	1	1	1	1	24	UTC	1	0,446	0,300	0,283	0,343	26%	0%	26%	10%	38%
1	1	1	1	1	1	24	UTC	2	0,256	0,181	0,161	0,200	22%	1%	8%	50%	19%
1	1	1	1	1	1	24	UTC	3	0,198	0,154	0,139	0,164	27%	2%	0%	62%	8%
1	1	1	1	1	1	24	UTC	4	0,171	0,110	0,069	0,117	57%	1%	0%	24%	19%
1	1	1	1	1	1	24	UTC	5	0,129	0,080	0,053	0,087	1%	22%	47%	28%	1%
1	1	1	1	1	1	24	UTC	6	0,129	0,061	0,041	0,077	3%	9%	39%	48%	1%
1	1	1	1	1	1	24	UTC	7	0,130	0,067	0,047	0,081	0%	51%	20%	21%	7%
1	1	1	1	1	1	24	UTC	8	0,157	0,092	0,076	0,108	28%	2%	20%	8%	43%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	1	1	24	UTC	9	0,133	0,079	0,057	0,090	33%	9%	11%	19%	28%
1	1	1	1	1	1	24	UTC	10	0,129	0,083	0,066	0,092	23%	33%	24%	0%	21%
1	1	1	1	1	1	24	UTC	11	0,125	0,077	0,064	0,089	20%	28%	27%	1%	25%
1	1	1	1	1	1	24	UTC	12	0,113	0,068	0,053	0,078	33%	30%	4%	3%	30%
1	1	1	1	1	1	24	UTC	13	0,108	0,069	0,054	0,077	28%	26%	3%	8%	36%
1	1	1	1	1	1	24	UTC	14	0,096	0,059	0,042	0,066	29%	38%	1%	1%	31%
1	1	1	1	1	1	24	UTC	15	0,088	0,049	0,037	0,058	34%	32%	6%	0%	28%
1	1	1	1	1	1	24	UTC	16	0,090	0,038	0,025	0,051	70%	1%	12%	1%	17%
1	1	1	1	1	1	24	UTC	17	0,074	0,044	0,030	0,049	81%	5%	5%	7%	2%
1	1	1	1	1	1	24	UTC	18	0,070	0,040	0,033	0,048	69%	24%	2%	4%	1%
1	1	1	1	1	1	24	UTC	19	0,064	0,036	0,029	0,043	71%	8%	2%	0%	19%
1	1	1	1	1	1	24	UTC	20	0,073	0,041	0,039	0,051	27%	43%	21%	3%	6%
1	1	1	1	1	1	24	UTC	25	0,078	0,030	0,025	0,044	16%	11%	24%	47%	3%
1	1	1	1	1	1	24	UTC	30	0,072	0,035	0,028	0,045	14%	5%	8%	70%	3%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	1	1	24	UTC	40	0,057	0,027	0,024	0,036	37%	32%	22%	6%	4%
1	1	1	1	1	1	24	UTC	50	0,057	0,026	0,023	0,035	24%	15%	9%	48%	3%
1	1	1	1	1	1	24	UTC	100	0,044	0,019	0,021	0,028	32%	38%	6%	18%	5%
1	1	1	1	1	1,5	24	CET	6	0,132	0,096	0,053	0,094	23%	3%	28%	0%	45%
1	1	1	1	1	1,5	24	CET	8	0,117	0,080	0,052	0,083	0%	12%	58%	17%	14%
1	1	1	1	1	1,5	24	CET	10	0,119	0,078	0,063	0,087	7%	17%	59%	0%	16%
1	1	1	1	1	1,5	24	CET	12	0,103	0,063	0,050	0,072	9%	1%	85%	1%	4%
1	1	1	1	1	1,5	24	CET	14	0,114	0,060	0,054	0,076	16%	3%	65%	15%	1%
1	1	1	1	1	2	24	CET	6	0,114	0,086	0,040	0,080	4%	39%	4%	1%	51%
1	1	1	1	1	2	24	CET	8	0,101	0,081	0,052	0,078	5%	69%	19%	7%	0%
1	1	1	1	1	2	24	CET	10	0,092	0,072	0,058	0,074	7%	40%	48%	5%	0%
1	1	1	1	1	2	24	CET	12	0,090	0,062	0,040	0,064	5%	80%	3%	12%	1%
1	1	1	1	1	2	24	CET	14	0,110	0,056	0,030	0,065	23%	54%	13%	8%	1%
1	1	1	1	1,5	1	24	CET	6	0,132	0,096	0,053	0,094	23%	3%	28%	0%	45%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	1	1,5	1	24	CET	8	0,117	0,080	0,052	0,083	0%	12%	58%	17%	14%
1	1	1	1	1,5	1	24	CET	10	0,119	0,078	0,063	0,087	7%	17%	59%	0%	16%
1	1	1	1	1,5	1	24	CET	12	0,103	0,063	0,050	0,072	9%	1%	85%	1%	4%
1	1	1	1	1,5	1	24	CET	14	0,114	0,060	0,054	0,076	16%	3%	65%	15%	1%
1	1	1	1	2	1	24	CET	6	0,124	0,092	0,049	0,088	3%	26%	5%	0%	65%
1	1	1	1	2	1	24	CET	8	0,105	0,079	0,045	0,076	7%	57%	26%	10%	0%
1	1	1	1	2	1	24	CET	10	0,091	0,067	0,048	0,069	10%	14%	68%	8%	0%
1	1	1	1	2	1	24	CET	12	0,091	0,056	0,019	0,055	20%	12%	12%	52%	3%
1	1	1	1	2	1	24	CET	14	0,112	0,053	0,024	0,063	35%	30%	20%	13%	2%
1	1	1	1,5	1	1	24	CET	6	0,145	0,090	0,072	0,102	24%	0%	21%	19%	36%
1	1	1	1,5	1	1	24	CET	8	0,118	0,076	0,076	0,090	14%	11%	7%	62%	6%
1	1	1	1,5	1	1	24	CET	10	0,108	0,065	0,068	0,081	15%	7%	43%	34%	2%
1	1	1	1,5	1	1	24	CET	12	0,106	0,051	0,053	0,070	12%	3%	47%	23%	15%
1	1	1	1,5	1	1	24	CET	14	0,095	0,061	0,065	0,074	19%	14%	59%	6%	3%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	1	2	1	1	24	CET	6	0,102	0,070	0,031	0,067	3%	22%	6%	65%	4%
1	1	1	2	1	1	24	CET	8	0,085	0,096	0,093	0,091	39%	8%	51%	0%	2%
1	1	1	2	1	1	24	CET	10	0,107	0,083	0,078	0,089	24%	6%	67%	3%	0%
1	1	1	2	1	1	24	CET	12	0,098	0,054	0,037	0,063	69%	1%	28%	2%	1%
1	1	1	2	1	1	24	CET	14	0,092	0,052	0,047	0,064	61%	3%	27%	4%	4%
1	1	1,5	1	1	1	24	CET	6	0,099	0,073	0,052	0,074	0%	4%	27%	4%	64%
1	1	1,5	1	1	1	24	CET	8	0,110	0,063	0,032	0,068	29%	5%	13%	20%	32%
1	1	1,5	1	1	1	24	CET	10	0,102	0,054	0,026	0,061	7%	71%	4%	11%	8%
1	1	1,5	1	1	1	24	CET	12	0,104	0,051	0,027	0,061	7%	50%	6%	4%	33%
1	1	1,5	1	1	1	24	CET	14	0,105	0,048	0,029	0,061	20%	25%	7%	27%	21%
1	1	2	1	1	1	24	CET	6	0,115	0,070	0,046	0,077	8%	1%	0%	3%	88%
1	1	2	1	1	1	24	CET	8	0,090	0,062	0,049	0,067	12%	12%	58%	12%	6%
1	1	2	1	1	1	24	CET	10	0,096	0,055	0,052	0,067	1%	1%	23%	58%	16%
1	1	2	1	1	1	24	CET	12	0,101	0,051	0,054	0,069	6%	0%	47%	39%	8%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1	2	1	1	1	24	CET	14	0,103	0,052	0,054	0,069	3%	6%	6%	62%	24%
1	1,2	1	1	1	1	24	CET	6	0,101	0,086	0,063	0,083	4%	26%	4%	49%	16%
1	1,2	1	1	1	1	24	CET	8	0,146	0,087	0,066	0,099	3%	4%	1%	36%	56%
1	1,2	1	1	1	1	24	CET	10	0,116	0,067	0,056	0,080	11%	0%	6%	2%	81%
1	1,2	1	1	1	1	24	CET	12	0,103	0,061	0,049	0,071	1%	1%	0%	32%	66%
1	1,2	1	1	1	1	24	CET	14	0,096	0,053	0,041	0,063	2%	3%	27%	16%	52%
1	1,2	1,2	1	1	1	24	CET	6	0,100	0,079	0,083	0,087	49%	1%	34%	9%	7%
1	1,2	1,2	1	1	1	24	CET	8	0,111	0,071	0,065	0,082	8%	18%	68%	4%	2%
1	1,2	1,2	1	1	1	24	CET	10	0,115	0,074	0,070	0,086	9%	6%	51%	0%	34%
1	1,2	1,2	1	1	1	24	CET	12	0,114	0,050	0,035	0,067	8%	2%	20%	5%	64%
1	1,2	1,2	1	1	1	24	CET	14	0,108	0,052	0,041	0,067	21%	13%	6%	44%	17%
1	1,3	1,3	1	1	1	24	CET	6	0,085	0,074	0,053	0,071	1%	0%	2%	82%	15%
1	1,3	1,3	1	1	1	24	CET	8	0,103	0,071	0,059	0,078	0%	22%	10%	51%	16%
1	1,3	1,3	1	1	1	24	CET	10	0,101	0,065	0,053	0,073	2%	15%	6%	29%	49%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1,3	1,3	1	1	1	24	CET	12	0,089	0,051	0,044	0,061	21%	11%	2%	49%	16%
1	1,3	1,3	1	1	1	24	CET	14	0,109	0,058	0,053	0,073	20%	21%	10%	35%	14%
1	1,4	1,4	1	1	1	24	CET	6	0,118	0,084	0,058	0,087	5%	34%	49%	12%	0%
1	1,4	1,4	1	1	1	24	CET	8	0,134	0,073	0,061	0,089	0%	21%	36%	36%	6%
1	1,4	1,4	1	1	1	24	CET	10	0,101	0,069	0,068	0,079	0%	8%	86%	5%	0%
1	1,4	1,4	1	1	1	24	CET	12	0,091	0,053	0,055	0,066	6%	8%	74%	4%	7%
1	1,4	1,4	1	1	1	24	CET	14	0,075	0,049	0,045	0,056	2%	18%	39%	18%	24%
1	1,5	1	1	1	1	24	CET	6	0,134	0,092	0,076	0,101	25%	2%	0%	16%	57%
1	1,5	1	1	1	1	24	CET	8	0,104	0,075	0,043	0,074	9%	25%	14%	0%	52%
1	1,5	1	1	1	1	24	CET	10	0,109	0,064	0,040	0,071	15%	6%	8%	4%	68%
1	1,5	1	1	1	1	24	CET	12	0,090	0,051	0,026	0,056	13%	12%	23%	2%	49%
1	1,5	1	1	1	1	24	CET	14	0,090	0,067	0,053	0,070	18%	28%	21%	10%	23%
1	1,5	1,5	0,5	0,5	0,5	24	CET	6	0,137	0,080	0,060	0,092	14%	24%	1%	27%	35%
1	1,5	1,5	0,5	0,5	0,5	24	CET	8	0,134	0,083	0,072	0,097	11%	0%	33%	45%	11%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1,5	1,5	0,5	0,5	0,5	24	CET	10	0,114	0,062	0,043	0,073	38%	48%	0%	8%	7%
1	1,5	1,5	0,5	0,5	0,5	24	CET	12	0,112	0,056	0,052	0,074	29%	1%	35%	29%	6%
1	1,5	1,5	0,5	0,5	0,5	24	CET	14	0,112	0,049	0,047	0,069	30%	3%	8%	59%	0%
1	1,5	1,5	1	1	1	1	CET	1	0,445	0,349	0,353	0,382	59%	12%	17%	11%	1%
1	1,5	1,5	1	1	1	1	CET	2	0,268	0,160	0,122	0,184	0%	6%	14%	58%	22%
1	1,5	1,5	1	1	1	1	CET	3	0,238	0,125	0,105	0,156	37%	3%	0%	49%	11%
1	1,5	1,5	1	1	1	1	CET	4	0,196	0,094	0,044	0,112	43%	6%	16%	8%	28%
1	1,5	1,5	1	1	1	1	CET	5	0,165	0,080	0,056	0,100	23%	0%	9%	5%	63%
1	1,5	1,5	1	1	1	1	CET	6	0,146	0,074	0,047	0,089	53%	12%	13%	0%	21%
1	1,5	1,5	1	1	1	1	CET	7	0,141	0,067	0,032	0,080	74%	11%	15%	1%	0%
1	1,5	1,5	1	1	1	1	CET	8	0,126	0,074	0,048	0,083	2%	56%	0%	38%	4%
1	1,5	1,5	1	1	1	1	CET	9	0,139	0,073	0,051	0,088	2%	46%	8%	35%	9%
1	1,5	1,5	1	1	1	1	CET	10	0,145	0,063	0,047	0,085	13%	47%	5%	30%	6%
1	1,5	1,5	1	1	1	1	CET	11	0,134	0,063	0,051	0,083	0%	57%	6%	24%	13%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1,5	1,5	1	1	1	1	CET	12	0,121	0,058	0,029	0,069	41%	4%	27%	17%	10%
1	1,5	1,5	1	1	1	1	CET	13	0,119	0,063	0,032	0,071	9%	1%	60%	26%	4%
1	1,5	1,5	1	1	1	1	CET	14	0,122	0,056	0,032	0,070	27%	12%	2%	60%	0%
1	1,5	1,5	1	1	1	1	CET	15	0,121	0,051	0,028	0,067	24%	16%	7%	53%	1%
1	1,5	1,5	1	1	1	1	CET	16	0,121	0,050	0,019	0,064	15%	44%	13%	11%	16%
1	1,5	1,5	1	1	1	1	CET	17	0,112	0,049	0,019	0,060	4%	59%	15%	0%	22%
1	1,5	1,5	1	1	1	1	CET	18	0,112	0,051	0,033	0,065	0%	48%	47%	1%	4%
1	1,5	1,5	1	1	1	1	CET	19	0,106	0,047	0,027	0,060	32%	44%	13%	9%	2%
1	1,5	1,5	1	1	1	1	CET	20	0,112	0,049	0,033	0,065	23%	16%	58%	3%	0%
1	1,5	1,5	1	1	1	1	CET	25	0,082	0,045	0,025	0,051	15%	0%	36%	45%	4%
1	1,5	1,5	1	1	1	1	CET	30	0,056	0,044	0,031	0,044	1%	3%	2%	34%	60%
1	1,5	1,5	1	1	1	1	CET	40	0,068	0,034	0,014	0,039	8%	9%	8%	34%	42%
1	1,5	1,5	1	1	1	1	CET	50	0,076	0,028	0,007	0,037	28%	19%	0%	27%	26%
1	1,5	1,5	1	1	1	1	CET	100	0,045	0,024	0,011	0,027	3%	52%	37%	0%	8%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1,5	1,5	1	1	1	24	CET	1	0,445	0,307	0,293	0,348	27%	0%	25%	12%	36%
1	1,5	1,5	1	1	1	24	CET	2	0,184	0,152	0,112	0,149	28%	23%	15%	3%	30%
1	1,5	1,5	1	1	1	24	CET	3	0,192	0,140	0,129	0,153	11%	43%	2%	8%	36%
1	1,5	1,5	1	1	1	24	CET	4	0,180	0,115	0,102	0,132	30%	32%	0%	1%	37%
1	1,5	1,5	1	1	1	24	CET	5	0,126	0,070	0,045	0,080	16%	4%	67%	13%	0%
1	1,5	1,5	1	1	1	24	CET	6	0,117	0,059	0,043	0,073	11%	7%	58%	25%	0%
1	1,5	1,5	1	1	1	24	CET	7	0,130	0,071	0,057	0,086	0%	0%	44%	25%	31%
1	1,5	1,5	1	1	1	24	CET	8	0,103	0,060	0,040	0,067	0%	1%	31%	51%	17%
1	1,5	1,5	1	1	1	24	CET	9	0,103	0,063	0,043	0,070	1%	0%	0%	83%	15%
1	1,5	1,5	1	1	1	24	CET	10	0,087	0,056	0,025	0,056	1%	9%	43%	46%	1%
1	1,5	1,5	1	1	1	24	CET	11	0,091	0,053	0,018	0,054	0%	3%	42%	54%	1%
1	1,5	1,5	1	1	1	24	CET	12	0,092	0,059	0,044	0,065	0%	1%	13%	9%	76%
1	1,5	1,5	1	1	1	24	CET	13	0,091	0,051	0,035	0,059	5%	0%	2%	44%	49%
1	1,5	1,5	1	1	1	24	CET	14	0,088	0,049	0,033	0,057	15%	10%	1%	39%	36%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1,5	1,5	1	1	1	24	CET	15	0,090	0,050	0,034	0,058	3%	16%	30%	20%	31%
1	1,5	1,5	1	1	1	24	CET	16	0,096	0,047	0,035	0,059	5%	16%	24%	16%	38%
1	1,5	1,5	1	1	1	24	CET	17	0,097	0,044	0,031	0,057	1%	17%	2%	19%	62%
1	1,5	1,5	1	1	1	24	CET	18	0,094	0,047	0,040	0,060	1%	37%	16%	2%	45%
1	1,5	1,5	1	1	1	24	CET	19	0,084	0,044	0,032	0,053	3%	83%	5%	2%	7%
1	1,5	1,5	1	1	1	24	CET	20	0,088	0,043	0,033	0,055	1%	43%	41%	3%	12%
1	1,5	1,5	1	1	1	24	CET	25	0,091	0,043	0,042	0,059	13%	39%	33%	0%	15%
1	1,5	1,5	1	1	1	24	CET	30	0,088	0,036	0,035	0,053	6%	35%	35%	5%	19%
1	1,5	1,5	1	1	1	24	CET	40	0,068	0,025	0,022	0,038	17%	12%	25%	15%	31%
1	1,5	1,5	1	1	1	24	CET	50	0,069	0,026	0,022	0,039	1%	9%	37%	16%	37%
1	1,5	1,5	1	1	1	24	CET	100	0,035	0,018	0,016	0,023	4%	35%	8%	33%	20%
1	1,6	1,6	1	1	1	24	CET	6	0,084	0,068	0,051	0,068	6%	46%	36%	11%	1%
1	1,6	1,6	1	1	1	24	CET	8	0,115	0,075	0,076	0,089	9%	38%	4%	25%	24%
1	1,6	1,6	1	1	1	24	CET	10	0,115	0,066	0,066	0,082	20%	46%	0%	22%	11%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	1,6	1,6	1	1	1	24	CET	12	0,097	0,059	0,050	0,069	19%	63%	3%	2%	13%
1	1,6	1,6	1	1	1	24	CET	14	0,096	0,037	0,026	0,053	29%	37%	18%	2%	14%
1	1,7	1,7	1	1	1	24	CET	6	0,096	0,073	0,062	0,077	3%	18%	28%	48%	2%
1	1,7	1,7	1	1	1	24	CET	8	0,136	0,071	0,054	0,087	4%	28%	13%	28%	27%
1	1,7	1,7	1	1	1	24	CET	10	0,118	0,060	0,048	0,075	17%	0%	70%	3%	9%
1	1,7	1,7	1	1	1	24	CET	12	0,128	0,068	0,063	0,086	13%	0%	69%	1%	17%
1	1,7	1,7	1	1	1	24	CET	14	0,109	0,061	0,065	0,078	20%	1%	73%	6%	0%
1	1,8	1,8	1	1	1	24	CET	6	0,088	0,070	0,064	0,074	5%	4%	68%	16%	6%
1	1,8	1,8	1	1	1	24	CET	8	0,135	0,082	0,082	0,099	0%	15%	22%	3%	60%
1	1,8	1,8	1	1	1	24	CET	10	0,137	0,080	0,086	0,101	4%	4%	44%	13%	34%
1	1,8	1,8	1	1	1	24	CET	12	0,123	0,067	0,072	0,087	8%	5%	49%	12%	26%
1	1,8	1,8	1	1	1	24	CET	14	0,107	0,051	0,045	0,068	22%	14%	51%	2%	11%
1	2	1	1	1	1	24	CET	6	0,093	0,062	0,042	0,066	3%	8%	77%	11%	2%
1	2	1	1	1	1	24	CET	8	0,119	0,083	0,065	0,089	0%	35%	32%	0%	32%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1	2	1	1	1	1	24	CET	10	0,122	0,084	0,076	0,094	31%	17%	32%	0%	20%
1	2	1	1	1	1	24	CET	12	0,116	0,074	0,056	0,082	28%	12%	42%	2%	15%
1	2	1	1	1	1	24	CET	14	0,112	0,068	0,058	0,079	43%	7%	6%	4%	39%
1	2	2	1	1	1	24	CET	6	0,092	0,086	0,078	0,085	0%	9%	84%	7%	0%
1	2	2	1	1	1	24	CET	8	0,093	0,068	0,048	0,070	6%	0%	49%	16%	29%
1	2	2	1	1	1	24	CET	10	0,113	0,051	0,038	0,067	26%	11%	42%	13%	8%
1	2	2	1	1	1	24	CET	12	0,113	0,070	0,079	0,087	15%	6%	74%	2%	4%
1	2	2	1	1	1	24	CET	14	0,106	0,067	0,074	0,082	6%	15%	78%	1%	0%
1,5	1	1	1	1	1	24	CET	6	0,146	0,102	0,058	0,102	6%	0%	5%	23%	66%
1,5	1	1	1	1	1	24	CET	8	0,126	0,079	0,061	0,088	8%	2%	3%	74%	12%
1,5	1	1	1	1	1	24	CET	10	0,095	0,053	0,029	0,059	23%	35%	0%	42%	0%
1,5	1	1	1	1	1	24	CET	12	0,102	0,052	0,041	0,065	31%	18%	18%	12%	21%
1,5	1	1	1	1	1	24	CET	14	0,106	0,054	0,048	0,069	48%	16%	3%	2%	30%
1,5	1,5	1,5	1,5	1	1	24	CET	6	0,111	0,092	0,038	0,080	22%	8%	11%	15%	44%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
1,5	1,5	1,5	1,5	1	1	24	CET	8	0,132	0,070	0,049	0,084	49%	3%	36%	1%	12%
1,5	1,5	1,5	1,5	1	1	24	CET	10	0,131	0,073	0,060	0,088	52%	4%	8%	2%	34%
1,5	1,5	1,5	1,5	1	1	24	CET	12	0,124	0,060	0,059	0,081	25%	23%	37%	6%	10%
1,5	1,5	1,5	1,5	1	1	24	CET	14	0,117	0,052	0,049	0,072	54%	1%	26%	8%	10%
2	1	1	1	1	1	24	CET	6	0,142	0,115	0,104	0,121	20%	1%	60%	7%	12%
2	1	1	1	1	1	24	CET	8	0,130	0,091	0,068	0,096	0%	1%	47%	30%	23%
2	1	1	1	1	1	24	CET	10	0,114	0,063	0,049	0,076	26%	2%	46%	8%	17%
2	1	1	1	1	1	24	CET	12	0,092	0,059	0,050	0,067	6%	13%	46%	2%	33%
2	1	1	1	1	1	24	CET	14	0,076	0,051	0,025	0,050	3%	0%	53%	5%	39%
2	2	2	1	1	1	24	CET	6	0,118	0,089	0,064	0,090	37%	0%	33%	2%	28%
2	2	2	1	1	1	24	CET	8	0,118	0,077	0,070	0,088	3%	8%	43%	2%	44%
2	2	2	1	1	1	24	CET	10	0,092	0,066	0,053	0,070	1%	0%	77%	5%	17%
2	2	2	1	1	1	24	CET	12	0,086	0,046	0,031	0,054	3%	1%	91%	4%	2%
2	2	2	1	1	1	24	CET	14	0,085	0,048	0,034	0,056	0%	17%	75%	8%	0%

Weight Load	Weight Solar	Weight Wind	Weight Run-of-river	Weight Pumped storage	Weight Reservoir	Downsampling (# hours)	Time	Clusters	NRMSD Price	NRMSD Generation	NRMSD Cost	Average of 3 metrics NRMSD	% NRMSD Cost 2016	% NRMSD Cost 2017	% NRMSD Cost 2018	% NRMSD Cost 2019	% NRMSD Cost 2020
2	2	2	2	1	1	24	CET	6	0,124	0,069	0,035	0,076	47%	13%	13%	9%	18%
2	2	2	2	1	1	24	CET	8	0,122	0,070	0,049	0,081	3%	3%	92%	0%	2%
2	2	2	2	1	1	24	CET	10	0,113	0,074	0,069	0,085	1%	0%	62%	3%	33%
2	2	2	2	1	1	24	CET	12	0,117	0,077	0,084	0,093	10%	13%	59%	0%	18%
2	2	2	2	1	1	24	CET	14	0,107	0,058	0,051	0,072	12%	12%	62%	14%	1%
3	1	1	1	1	1	24	CET	6	0,124	0,121	0,093	0,113	44%	0%	36%	4%	16%
3	1	1	1	1	1	24	CET	8	0,106	0,059	0,039	0,068	12%	78%	4%	5%	1%
3	1	1	1	1	1	24	CET	10	0,122	0,094	0,068	0,094	15%	54%	2%	18%	10%
3	1	1	1	1	1	24	CET	12	0,092	0,070	0,059	0,074	18%	16%	9%	37%	19%
3	1	1	1	1	1	24	CET	14	0,093	0,060	0,048	0,067	29%	20%	1%	21%	28%