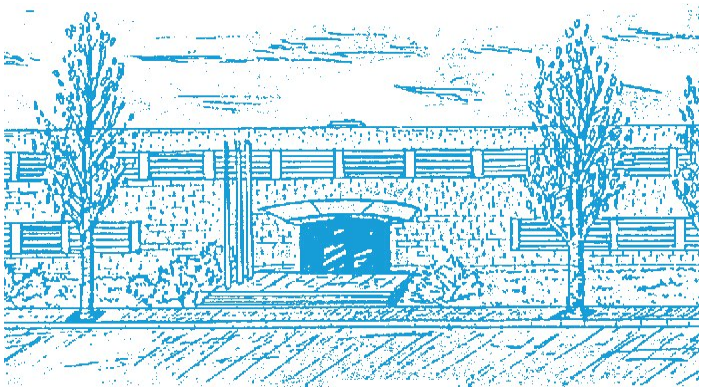
Master of Science in Advanced Mathematics and Mathematical Engineering

Title: Forecasting Country-Level Electricity Demand Using Temperature Data and a Clustering Approach in Constructing Sub-Country Level Prediction Models

Author: Carl Ian Wallgren
Advisor UPC: Antonio Agudo
Advisors BSC: Albert Soret & Verónica Torralba
Department: Facultat de Matemàtiques i Estadística
Academic year: 2024



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Forecasting Country-Level Electricity Demand Using Temperature Data and a Clustering Approach in Constructing Sub-Country Level Prediction Models

Dissertation: Master of Science in Advanced Mathematics and Mathematical Engineering

Author

Carl Ian Wallgren

Supervisors Barcelona Supercomputing Center:

Albert Soret and Verónica Torralba

Supervisor Universitat Politècnica de Catalunya:

Antonio Agudo

June 18, 2024

Acknowledgements

This thesis would not have been possible to complete without the collaboration with Barcelona Supercomputing Center (BSC). BSC's computational resources and environment of excellent researchers have been imperative in facilitating the progression of this dissertation.

Specifically, I owe special thanks to a few members of the BSC. Firstly, I would like to extend a sincere thank you to my supervisors, Dr. Albert Soret and Dr. Verónica Torralba, for acknowledgeing this research opportunity to begin with, and for generously lending me their time throughout the process. I am also grateful for the fruitful discussions had with Dr. Angel Garikoitz Munoz; these have been inspiring and essential in finding solutions to obstacles encountered in the research. Additionally, I want to express my gratitude to BSC PhD student Lluís Palma Garcia for being patient with my many questions and for his technical support all through this study.

Further, I am truly thankful for the indispensable support I have received from my supervisor at the Universitat Politècnica de Catalunya (UPC), Antonio Agudo. Your expertise in the field of machine learning and genuine interest in this dissertation have been vital in transforming this research process into a successful study as well as a great learning experience.

Abstract

As the battle against climate change progresses, a transition to more renewable energy sources in the mix of global energy is vital. By 2016, the consumption of the EU's energy was to more than 50 percent allocated towards heating and cooling, essentially making electricity demand strongly dependent on prevailing weather conditions. This has laid the foundation for a well-studied area within the climate science community: forecasting electricity demand as a function of temperature data, in which common tools are the so-called degree-day variables, aspiring to translate temperature data to an indicator of electricity needed for the purpose of heating and cooling.

The purpose of this dissertation is to build upon currently established and successful methodologies in the prediction of electricity demand, aiming to combine the most promising research findings within the field and elaborate on these by the inclusion of widely-used and effective clustering methodologies from the context of machine learning. Particularly, with the starting point of an existing electricity demand prediction model, the intention of this research is two-fold; to construct and assess the performances of a wide range of alterations to conventional degree-day methods, and to explore the opportunities in developing regional models without access to higher than country-level resolution of observed electricity demand, with the aim of increasing model accuracy and robustness by accounting for the heterogeneous climate of the country for which electricity demand is predicted. Specifically, this paper will focus on predicting electricity demand in the Spanish Peninsula.

Contents

1 Introduction

In the battle against climate change, the transition to more renewable energy sources in the global energy mix is imperative. With a sustained route route toward electrification, the world will in the next five years have added more renewable energy capacity than what has been incorporated between today and a 100 years ago, when the very first commercial renewable energy power plant was constructed [IEA, 2O23]. Out of the total energy consumption, electricity constitutes approximately 20 percent, a number which is expected to increase to 50 percent by 2050 as an effect of decarbonisation [IEA, 2O23]. The supply of such electricity to homes and businesses is in most countries where the electricity market is liberalized managed by independent Transmission System Operators (TSOs) [Caro et al., 2019]. Due to the difficulty and cost of storing energy [Gates, 2016], such large-scale power system operators are met with a difficult task of ensuring balance between electricity generation and demand on a near-instantaneous basis [Bloomfield et al., 2021]. Thus, accurate forecasts of demand have been pivotal in the power industry [Hong et al., 2020]. In-advance knowledge of how electricity demand will fluctuate helps streamline the operation of the power system, improving efficiency and reliability of energy management at both national and local scales [Bloomfield et al., 2021].

1.1 Modeling electricity demand as a function of weather data

The heating and cooling consumption corresponds to half of the EU's energy [Commission, 2016], giving rise to strongly weather-sensitive components in electricity demand [Bloomfield et al., 2021]. Modeling the electricity demand as a function of weather data is a well-studied topic, in which the use of the concept of degree-days is a common choice [Thom, 1954]. In layman's terms, degree-days aspire to provide a simple measure of how warm or cold a location is by comparing a selected base-temperature to the current recorded temperature of the location. The more extreme the temperature is, in either direction, the higher is the number of the degree-day variable [EIA, 2023]. For low temperatures, degree-days increase due to the need for space heating, whereas, conversely, high temperatures increase degree-days due to the need for space cooling [Antunes-Azevedo et al., 2015].

Among a myriad of application areas, ranging from benchmarking energy efficiency of buildings [Chung, 2012] to projecting potential future climate change scenarios [Spinoni et al., 2017], one common use-case of degree-days is divided into two standalone indicator variables, namely cooling degree-days (CDD) and heating degree-days (HDD), aiming to provide a measure capturing the energy required to cool and heat buildings, respectively, according to predefined temperature thresholds [Kheiri et al., 2023].

Condition	HDD
$T_{max} \leq T_{base_{HDD}}$	$T_{base_{HDD}} - T_{avg}$
$T_{avg} \leq T_{base_{HDD}} \leq T_{max}$	$\frac{T_{base_{HDD}} - T_{min}}{2} - \frac{T_{max} - T_{base_{HDD}}}{4}$
$T_{min} < T_{base_{HDD}} < T_{avg}$	$\frac{T_{base_{HDD}} - T_{min}}{4}$
$T_{min} > T_{base_{HDD}}$	0

Table 1: **Calculation Heating Degree-Days (HDD).** Four intervals are considered as a consequence of applying different conditions [Spinoni et al., 2017].

However, there exists multiple methods of calculating degree-days with and which base-temperature to use can vary depending on, for instance, thermal capacity of buildings [Kheiri et al., 2023]. In Tables 1-2, the method that was developed by the UK MET Office in 1928 is depicted, which calculates degree-days by using $T_{avg} = \frac{T_{max} - T_{min}}{2}$, with T_{min} and T_{max} the daily minimum and maximum temperatures, respectively. $T_{base_{CDD}}$ and $T_{base_{HDD}}$ are predefined base-temperatures for cooling (usually set to 22 degrees) and heating (usually set to 15.5 degrees) degree-days, respectively. As is visible in the Tables, the calculations of the degree-day variables correspond to four different cases: 1) For $T_{max} \leq T_{base_{HDD}} < T_{base_{CDD}}$ the temperature on the day is consistently cold, yielding a need for heating and no need for cooling. 2) $T_{avg} \leq T_{base_{HDD}} < T_{base_{CDD}} \leq T_{max}$ corresponds to a day of predominantly cold temperature, resulting in a combination of heating and cooling to achieve a comfortable indoor temperature (with heating still more significant than cooling). 3) conversely, $T_{min} < T_{base_{HDD}} < T_{base_{CDD}} < T_{avg}$, corresponds to a day of primarily warm temperature, meaning a combination of heating and cooling is required but with cooling more significant than heating. 4) $T_{min} > T_{base_{CDD}} > T_{base_{HDD}}$ translates to a scenario where temperature is uniformly warm, thus no heating is required whereas cooling is [Spinoni et al., 2017]. The interval in-between the two thresholds is obviously 0 for both HDD and CDD; the range for which no heating or cooling is needed in order to attain a comfortable indoor temperature [Copernicus, 2022].

1.1.1 Contribution of this thesis

With a call for increased collaboration across energy and meteorological research communities, the authors in [Bloomfield et al., 2021] have helped lay the foundation for this thesis. The aim of this work is to further build on essential ideas presented by [Bloomfield et al., 2021], yet elaborate and adapt such further by drawing inspiration from previously mentioned successful strategies in the realm of electricity

Condition	CDD
$T_{max} \leq T_{base_{CDD}}$	0
$T_{avg} \leq T_{base_{CDD}} \leq T_{max}$	$\frac{T_{max} - T_{base_{CDD}}}{4}$
$T_{min} < T_{base_{CDD}} < T_{avg}$	$\frac{T_{max} - T_{base_{CDD}}}{2} - \frac{T_{base_{CDD}} - T_{min}}{4}$
$T_{min} > T_{base_{CDD}}$	$T_{avg} - T_{base_{CDD}}$

Table 2: **Calculation Cooling Degree-Days (CDD).** Four intervals are considered as a consequence of applying different conditions [Spinoni et al., 2017].

forecasting, applying and evaluating them for a case study of Spain. We think, however, that our approach could be easily applied for other countries as well.

In particular, when predicting the electricity demand for Peninsular Spain's electricity system, this thesis will focus on exploring the potential in terms of enhanced model accuracy and robustness by applying one or more of the following concepts to the main idea as presented in [Bloomfield et al., 2021]:

- A novel split degree-day approach to calculate degree-days, potentially entailing a decreased sensitivity to selected baseline temperatures.
- A population-weighted degree-days approach, aiming to assign greater significance to the temperature of more densely populated areas of the country.
- Replacing the country-level model by multiple local models by the means of:
  - An arbitrary country-split in terms of compass direction (North-West, North-East, South-West, South-East).
  - A k-means clustering approach.

The structure of this thesis will be as follows: firstly, a theoretical review of related work to the ideas presented above shall be presented, concluding with the current state-of-the-art model as presented in [Bloomfield et al., 2021]. Thereafter, the baseline formulation shall be presented, in terms of the previously derived state-of-the-art model and an in-depth description of the data being used. Following that, the main ideas listed above shall be outlined and explained further, leading to the final set of constructed models that will be tested and compared to the baseline model. After that, results will be demonstrated and analysed, concluding with some final remarks including potential insights that can be drawn from this thesis and brief notes on ideas worth exploring further.

2 Related work

As it was pointed out by [Antunes-Azevedo et al., 2015], when calculating degree-days, the definition of thresholds based on ambient or outdoor air temperature, such as in Tables 1-2, as opposed to the actual indoor temperature, will differ across countries, and even within countries. The authors also indicate the heterogeneity in the space of base-temperatures for calculating degree-days to be arguably the biggest limitation of the method. This is further highlighted in [Shi et al., 2016], where the variance of base-temperatures is attributed to the level of economic development of the country, climatic conditions, general statistics of buildings, etc. The findings in [Kennard et al., 2022] strengthened this hypothesis, concluding the parameter selection for the base-temperatures to be the single most important feature in terms of variability of degree-days.

Another feature subject to heterogeneity amongst research articles addressing electricity demand forecasting is that of population-weighting. An early adoption of this concept is seen in [Quayle and Diaz, 1980], where it is claimed that (heating) degree-days ought to be weighted by population in order to more adequately reflect fuel demand. As is explained in [EIA, 2023], when calculating population-weighted degree-days in the U.S., one approach has been to divide the country into nine Census regions for which weights are assigned on basis of the population count in the region relative to the population count of the nation as a whole. Thereafter, degree-day recordings are weighted by the corresponding population weight, essentially attributing greater importance to more densely populated regions. The concept of population-weighted degree-days was further utilised in [Taylor, 1981] as a mean to attain realistic estimates of potential fuel demand for Canada. In a more recent study, [Kennard et al., 2022] demonstrated the importance of considering the population weight when calculating degree-days, arriving at a statistical significant difference in the global rate of change between area-weighted cooling degree-days and population-weighted cooling degree-days.

In a case study of Spain, [Valor et al., 2001] collected weather data from four weather stations that were deemed to be representative of peninsular Spain as well as geographically corresponding population statistics. Thereafter, a population-weighted temperature index representing the most dominant areas of electricity consumption in Spain was constructed, which when regressed on actual consumption data showed a correlation coefficient between $0.79 - 0.86$. In the same paper, it was shown that electricity load is relatively insensitive to changes in air temperature around an interval centered at 18 degrees Celsius.

Just as there is heterogeneity in the space of possible base-temperature thresholds, there exist a variety of methods with which degree-days can be calculated. Apart from the MET expressions in Tables 1 and 2, another approach to calculate degree-days is presented in [Bloomfield et al., 2021] by simply using the mean temperature of the day and comparing it to the predefined base-temperatures, as will be shown in more detail later.

Another noteworthy observation made by [Kheiri et al., 2023] came as the authors designed a novel approach to the traditional methods of calculating degree-days; opposed to using base-temperature sensitive degree-days (as highlighted previously) as a measure of building energy use, the authors proposed the so-called *split degree-day* method. The method builds on partial degree-days calculated for two intervals of each day: one covering peak electricity demand hours, and the other one covering the remaining hours of the day. It could be shown that the split degree-day method performed superior to conventional ways of calculating degree-days, and more notably was much less sensitive to the selected base-temperature. The authors argue that, when estimating building energy consumption, conventional degree-days yield significant inaccuracies. Parameters that essentially determine the electricity consumption (e.g., efficiency of a fan) might vary substantially throughout a day, leading to imprecision as data is aggregated to a daily basis. Thus, by considering more information associated with the weather characteristics, as opposed to limiting this information to the daily mean temperature, for instance, such inaccuracies could be counteracted. By separating peak-hours from the rest of the hours of the day, [Kheiri et al., 2023] makes a strong case when arguing that such an approach more accurately captures daily fluctuations of temperature. When considering temperature data from 801 different locations in the U.S., the authors proceed to show the superiority of the split degree-day method in relation to a conventional benchmark model with an improvement of more than 5 percent in the accuracy of estimated total annual energy usage, followed by an 8 percent improvement in terms of the corresponding heating energy usage (forecast horizon unknown).

In symbiosis with a new EU initiative *Subseasonal-to-seasonal forecasting for Energy* (S2S4E), Bloomfield et al. [2021] presents a variety of different country-level, daily indicators for 28 European countries, aiming to provide a fully calibrated, post-processed power system forecast as the first one of its kind. In addition to skillful indicators covering wind power and solar photovoltaic power generation, the article presents a multiple linear regression model based on the concept of degree-days, in-

tended to capture weather-dependent variations of nationally aggregated electricity demand. The multiple linear regression model is trained separately for each of the 28 countries on observed country-level daily total electrcity load from 2016-2017, retrieved from the ENTSO-E transparency platform, and thereafter validated on data from 2018 [Bloomfield et al., 2020]. The model is presented in detail in Section 3.

3 Baseline

The baseline model for this research will be the one presented in [Bloomfield et al., 2021]. The authors use a model that maps temperature data to an estimate of nationally aggregated electricity load. This multiple linear regression model is given by:

$$Demand(t) = \alpha_0 + \alpha_1(t) + \alpha_2 HDD(t) + \alpha_3 CDD(t) + \sum_{i=4}^{9} \alpha_i DAY(t), \quad (1)$$

where, as in [Bloomfield et al., 2021], t is the time step in days (in [Bloomfield et al., 2021], count starts at 01/01/2016). In the model, α_0 and α_1 are terms aiming to capture the constant background level of demand, whereas α_4 to α_9 are dummy variables capturing the day-of-week effect (the 7-th dummy variable is as by standard not included due to multicollinearity issues). Lastly, α_2 and α_3 are intended to capture the sensitivity of electricity as a function of heating and cooling degree-days, respectively, with:

$$HDD(t) = \begin{cases} 15.5 - T(t) & \text{if } T(t) < 15.5 \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

$$CDD(t) = \begin{cases} T(t) - 22 & \text{if } T(t) > 22 \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

where T denotes the daily mean of the country-average temperature. It is highlighted by the authors that human behavioral factors, as captured by α_1 and α_{4-9}, might be removed, yielding an estimation of merely the weather-dependent demand. The full model, however, achieves an average R2-score of 0.80 [Bloomfield et al., 2021]. Before reviewing further the concept of the conventional degree-day method used in this baseline model (demonstrated in Eqs. (2) (3)), the data used, both for the baseline model as well as alternative models presented later in this report, shall be presented.

3.1 Data

3.1.1 Temperature data

The temperature data is retrieved from the ERA5-land reanalysis data set, with a temporal resolution of 1-hour and spatial grid resolution of 0.1 degrees (around 9 km in both latitudinal and longitudinal direction). The reanalysis data set reconstructs the recent atmosphere by combining model data with observations and provides accurate

descriptions of the climate, going back multiple decades [Copernicus, 2024a]. From the reanalysis dataset, 2m air temperature is retrieved from which the corresponding degree-day indicator value is calculated. It should be mentioned that the resolution of 0.1 degrees is higher than that used in [Bloomfield et al., 2021], however, their choice of reducing the spatial resolution to 1.5 degrees was associated with the need for the temperature data to be compatible with the resolution of used forecasting models. Further, considering the given setup of the baseline model, it is necessary for this purpose to have the data resampled to daily resolution.

3.1.2 Electricity demand data

The electricity data used in this study is observed national electricity demand with hourly temporal resolution and is obtained through the ENTSO-E Transparency Portal. The available history of such data for Peninsular Spain is from 2014-12-19 to present day, however, the most recent data is provided in 15-min resolution and has therefore been re-sampled to 1-hourly resolution to be compatible with the rest of the data. For the baseline model, this data will be further re-sampled to daily temporal resolution in order to be compatible with the format of the temperature data in the multiple linear regression model.

3.1.3 Data visualization

In order to understand the construction of the baseline model, it is necessary to review the main characteristics of the temperature as well as electricity load in Spain.

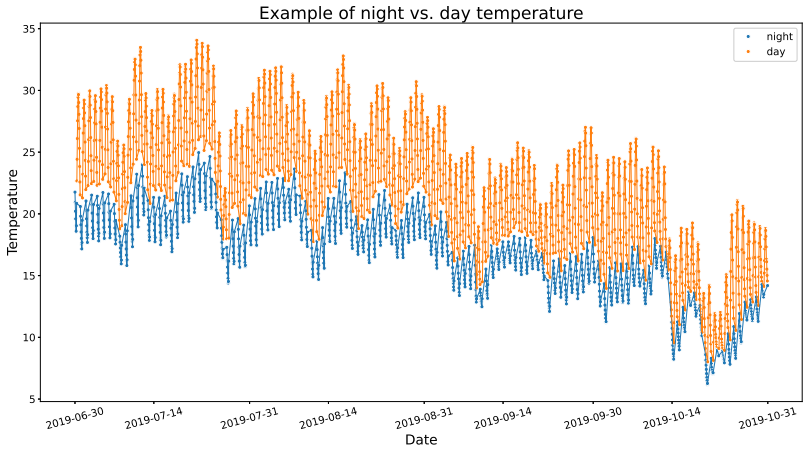


Figure 1: **Example plot of temperature with difference between day and night** highlighted (data points have been arbitrarily classified as to belong to the class "night" if the recorded hour falls in-between 23pm-7am).

As is clearly demonstrated in Fig. 1, temperature in Spain follows an annual cyclical variation; temperature troughs during winter, increases throughout spring and peaks during the summer months, and thereafter gradually sinks during the months of fall. It is also obvious that, throughout the year, day temperatures are generally higher than night temperatures. Thus, there is an additional 24-hour cyclical variation pertaining to temperature. A measure to avoid this daily variation, as is done by [Bloomfield et al., 2021], is to re-sample the data to daily format.

Due to the dependence on temperature for electricity for heating and cooling, with temperature as previously stated displaying distinct annual cyclical variations, electricity load exhibits a bi-annual cyclical variation; as temperature troughs, electricity demand for heating peaks, and as temperature peaks, electricity demand for cooling also peaks. Hence, there is generally a strong positive correlation between the two variables during summer months, followed by a negative correlation during winter months. Considering the construction of the HDD and CDD functions (see Eqs. (2)-(3)), both degree-day variables will be correlated with temperature as follows:

- in winter-time, temperature decreases → there is a need for heating → electricity demand increases (peak in electricity load).
- in spring-time, temperature increases → need for heating drops (and there is yet relatively little need for cooling) → electricity demand drops (trough in electricity load).
- in summer-time, temperature increases → need for cooling increases → electricity demand increases (peak in electricity load).
- in fall-time, temperature decreases → need for cooling drops (and there is yet relatively little need for heating) → electricity demand drops (trough in electricity load).

Further, electricity load displays a daily cyclical variation as demand troughs during night-time and peaks during day-time. However, considering that both temperature and electricity demand variables are re-sampled to daily resolution, such cyclical variation can be disregarded.

Lastly, by observing Fig. 2, it is clear that there is also a weekly cyclical variation in electricity demand; indeed, electricity consumption normally drops during weekends. Hence, it stands to reason to include α_{4-9} in Eq. (1), aiming to capture this weekly variation by the means of dummy variables.

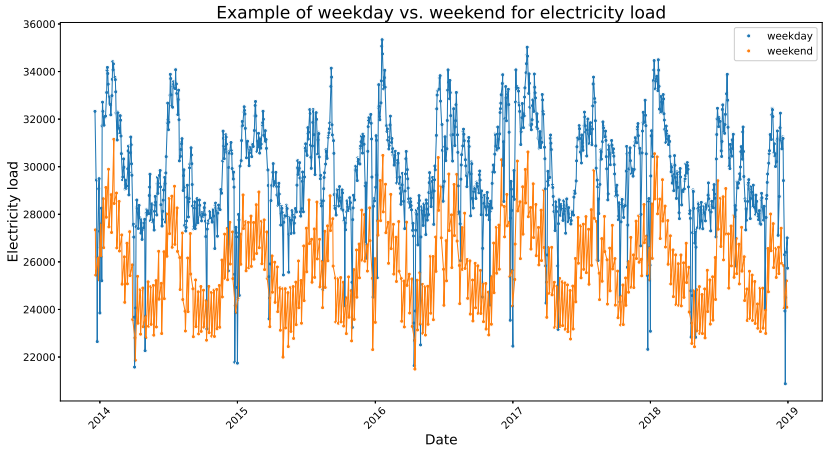


Figure 2: **Example plot of load**, highlighting the difference between weekday and weekend.

3.2 Conventional degree-day method

As previously described, there is a plethora of possible methods one can utilise when calculating the degree-day variables. This baseline method of calculating HDD and CDD will therefore be limited to the one presented in [Bloomfield et al., 2021] as shown in Eqs. (2)-(3). As previously stated, the daily-average country-mean is used as input to both the HDD and CDD functions and is approximated by considering all ERA5 reanalysis grid boxes (with a spatial resolution of 0.1 degrees) using a country mask and averaging across these. Applying a country mask corresponding to Spain allows for a selection of temperature values recorded only for grid boxes included within the specified region. Thus, as is done in [Bloomfield et al., 2021], country-average temperature for each point in time (daily, in this case) can be calculated as follows:

$$T_{country} = \frac{1}{|S|} \sum_{s \in S}^{|S|} T_s, \tag{4}$$

where S denotes the family of grid boxes within the borders of the specified country mask, and T denotes the average daily temperature for a specific grid box.

It should be noted that, as was previously discussed, the thresholds of 15.5 and 22 degrees Celsius for the HDD and CDD variables, respectively, could be subject to change depending on context in and location where the model is deployed. However, these thresholds are utilised for all 28 countries in [Bloomfield et al., 2021], and will therefore be applied to the baseline model used in this paper as well. Figure 3 demonstrates the relationship between the degree-day variables and country-average

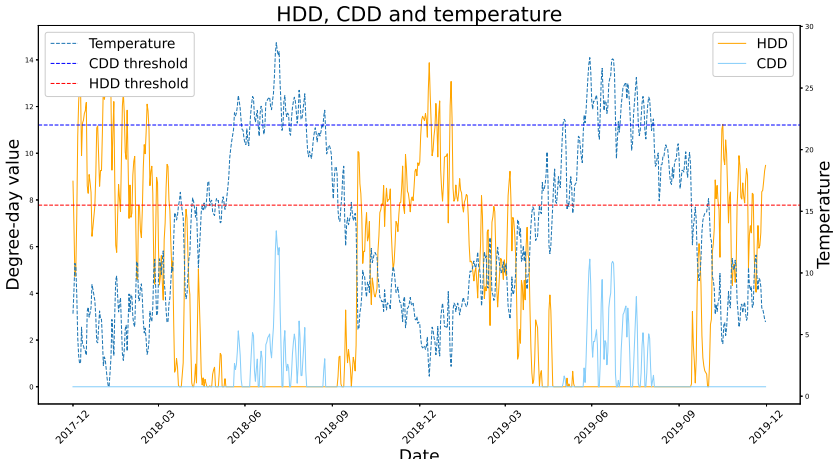


Figure 3: **Example plot of degree-day variables** and their relation to temperature based on "standard" (22 and 15.5 degrees Celsius for CDD and HDD, respectively) base-temperatures.

daily-mean temperature; for temperature values lying below the HDD threshold (the lower threshold corresponding to 15.5 degrees Celsius), it is assumed there is a need for heating, thus the HDD indicator assumes positive values whereas CDD indicator is 0, and for values of temperature above the CDD threshold (the upper threshold corresponding to 22 degrees Celsius), the assumption is that cooling is required, and the CDD indicator therefore takes positive values whereas HDD is identically 0.

4 Methodology

4.1 Model score measurements

Before going deeper into the methodology of this research, the two score measurements that will be consistently used for the evaluation of each model presented herein shall be stated. The first measurement, commonly referred to as the coefficient of determination, measures the proportion of the variance of the observed data explained by the model and is given by:

$$R2_{score} = 1 - \frac{SS_{res}}{SS_{tot}}, \tag{5}$$

where $SS_{res} = \sum_{i=1}^{n}(y_i - \tilde{y_i})^2$ and $SS_{tot} = \sum_{i=1}^{n}(y_i - \hat{y})^2$, with $y_i, \tilde{y_i}$ and $\hat{y}$ being the actual value of the dependent variable, the predicted value of the dependent variable from the regression model, and the mean of the observed data, respectively. Secondly, the Mean Absolute Percentage Error (MAPE) score will be utilised, which measures the average error size of predictions, and it is given by:

$$MAPE_{score} = 100 \cdot \frac{1}{n} \sum_{i=1}^{n} |\frac{y_i - \tilde{y_i}}{y_i}|, \tag{6}$$

with the variables consistent with the variables in Eq. (5).

4.2 Novel approach: Split degree-days

Initially, the strategy was to construct many local models (one for each region in Peninsular Spain) using regional electricity demand data, and aggregate the corresponding predictions to a global, country-level prediction. This would enable the application of locally optimized models using locally optimal thresholds for the degree-days variables. However, due to data restrictions, such approach was not feasible and country-level data had to be used instead. Therefore, it was desired to extend the methodology of degree-day calculations as used in [Bloomfield et al., 2021], with the aspiration to investigate whether it would also be possible to construct a method more robust and less sensitive to the specific base thresholds being used.

To this end, the alternative, novel method of the split degree-day calculation as presented in [Kheiri et al., 2023], with the property of being potentially less sensitive to the selected base-temperature thresholds, shall be presented more in-depth. The split degree-day indicator, as opposed to the conventional degree-day method described previously, constitutes of two partial degree-day variables that are derived separately for two disjoint time intervals in a day. The authors denote the interval in which the peak temperature of the day occurs as (v, w), where $1 \leq v$ and

$w \leq 24$, and the remaining two intervals (hours before and after the peak interval) as $(1, v') \cup (w', 24) = (v, w)'$ with $v' \leq v$, $w \leq w'$. Thereafter, the split degree-day indicators sHDD and sCDD are calculated for both intervals as follows:

$$sHDD_{(v,w)} = \left(T_{base_{HDD}} - \frac{T_{(v,w)_{max}} + T_{(v,w)_{min}}}{2} \right)^{+}, \tag{7}$$

$$sHDD_{(v,w)'} = \left(T_{base_{HDD}} - \frac{T_{(v,w)'_{max}} + T_{(v,w)'_{min}}}{2} \right)^{+}, \tag{8}$$

and

$$sCDD_{(v,w)} = (\frac{T_{(v,w)_{max}} + T_{(v,w)_{min}}}{2} - T_{base_{CDD}})^{+}, \tag{9}$$

$$sCDD_{(v,w)'} = (\frac{T_{(v,w)'_{max}} + T_{(v,w)'_{min}}}{2} - T_{base_{CDD}})^{+}, \tag{10}$$

with + indicating only non-negative values are considered.

At this point, the reader might appreciate an example of an arbitrary configuration. If the split-hour (s) corresponds to 13:00 and the window-size (δ) is 3 hours, the two intervals are defined as follows:

$$(v, w) = (s - \delta, s + \delta) = (13 - 3, 13 + 3) = (10, 16), \tag{11}$$

$$(v, w)' = (1, v - 1) \cup (w + 1, 24) = (1, v') \cup (w', 24) = (1, 9) \cup (17, 24). \tag{12}$$

Thereafter, optimal split-hour (split-hour := median hour [= mean hour due to symmetry] of $(v, w) = \frac{v+w}{2}$) and window-size (i.e. $|\frac{v+w}{2} - v'| - 1 = |\frac{v+w}{2} - w'| - 1$) are derived by considering (possible split-hours) $\cdot$ (possible window-sizes) $= 24 \cdot 12 = 288$ different configurations. For finding such optimal combination, [Kheiri et al., 2023] considered the coefficients of determination (R2-score) of simulation models estimating the yearly energy consumption of buildings, using base-temperatures $T_{base_{HDD}} = 18$ and $T_{base_{CDD}} = 10$. Eqs. 7-(10) are visualised in Fig. 4, where a pair of arbitrary split-hour and window-size has been chosen for both degree-day variables.

As in [Kheiri et al., 2023], different combinations of peak-hours and window-sizes are tested, with, as previously stated, the objective of finding a configuration that is relatively insensitive to the base-temperature thresholds for the degree-day variables inputted, while still achieving a relatively high measure-of-fit. In order to limit the computational load, the range of peak-hours and window-sizes will be limited to the

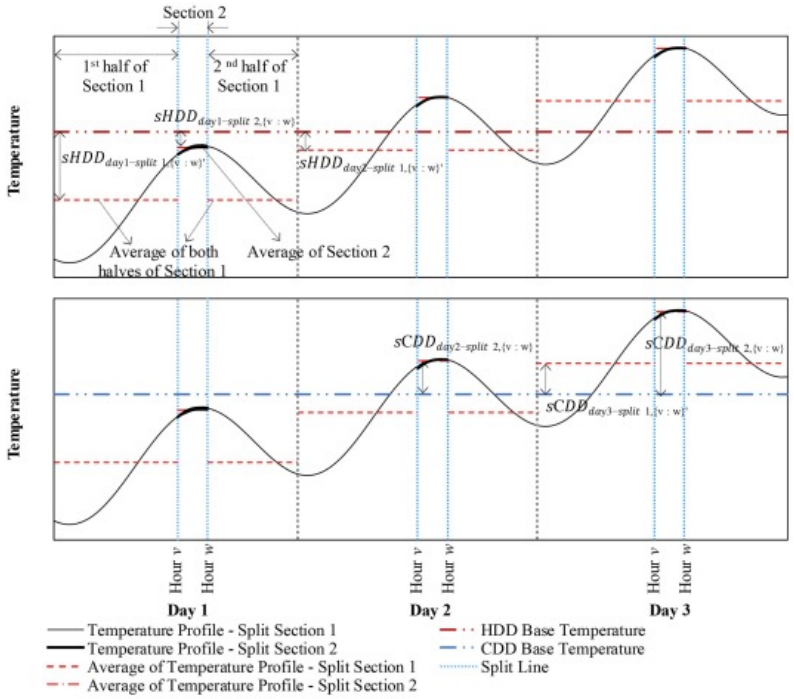


Figure 4: Illustration of novel split degree-day method. Image taken from [Kheiri et al., 2023].

following sets:

$$peak - hours = \{8am, 9am, 10am, 11am, 12am, 1pm, 2pm, 3pm, 4pm, 5pm\},$$
$$window - sizes = \{1hr, 2hr, 3hr, 4hr, 5hr, 6hr\},$$

yielding a total of |peak-hours|·|window-sizes| $= 10 \cdot 6 = 60$ combinations. It shall be emphasized again that [Kheiri et al., 2023] experimented with a total of |set of possible peak-hours| · |set of possible window-sizes| $= 24 \cdot 12 = 288$ combinations, however, based on the underlying hypothesis of the split degree-day concept, the 60 combinations derived previously are deemed sufficient (regardless of the time of the year, the peak temperature of the day ought to be covered by at least one of the intervals tested for, which is the intrinsic idea of the split degree-day method).

For each of these combinations, the multiple linear regression model given in Eq. (1) is trained and evaluated using a 10-fold Cross-Validation (CV), using country-level electricity demand data for all possible combinations of CDD and HDD threshold parameters. For each combination, the resulting R2-score is the average of all 10 R2-scores calculated in the 10-fold CV. Again, to assure a feasible computational load, the cardinality of the two sets of threshold parameters shall for this purpose be

limited to 5:

- CDD = $\{15.0, 17.5, 20.0, 22.5, 25.0\}$,
- HDD = $\{10.0, 12.5, 15.0, 17.5, 20.0\}$.

These two sets are deemed representative of probable, "true" thresholds for both HDD and CDD variables in any region of Peninsular Spain. Hence, for every combination of peak-hour and window-size, there will be $|\text{CDD}| \cdot |\text{HDD}| = 5 \cdot 5 = 25$ total combinations of CDD and HDD threshold parameters evaluated. In order to obtain a split-hour and window-size configuration yielding a relatively high measure-of-fit while simultaneously considering the robustness of the model (i.e., how sensitive the model is to changes in the base thresholds for HDD and CDD), a signal-to-noise ratio is for each configuration calculated as follows:

$$signal_{(peak-hour, window-size)} = \frac{median_{score}}{IQR_{score}} = \frac{median_{score}}{Q_{75} - Q_{25}}. \tag{13}$$

Finally, the combination of peak-hour and window-size that produces the highest signal value shall be considered as the optimal configuration for the novel split degree-day calculation in this setting.

4.3 Population-weighting

As previously brought up, another potential source of increased predictability of the baseline model formulation is the approach of weighting the degree-day values by the corresponding population density for each grid box. Hence, an in-depth description of what type of available population data to be included and how it will be applied will follow.

4.3.1 Population data

The population data is made available as Global Human Settlement (GHS) Population variable by Copernicus and comes with several spatial resolutions as well as Coordinate Reference Systems (CRS) in 5-year intervals (starting from 1975). The raster (grid) contains information of the human settlement and is expressed as the number of people per grid cell [Copernicus, 2024b]. For the purpose of this research, population data with a spatial resolution of 30 arcsec $\approx 31m$ and WGS84 CRS (i.e., the same CRS as temperature data is expressed in) is retrieved. For the population data to be compatible with the temperature data from the ERA5-land data set (and thus the temperature-derived degree-day values of the same resolution), a spatial aggregation with final spatial resolution of 0.1 degrees is realized.

In order to be able weight degree-day values per grid cell, population count data is converted to population density data; each grid cell value is expressed as relative to the total count of the whole grid covering Peninsular Spain, calculated as follows:

$$popdensity_p = \frac{count_p}{\sum_{count_p, p \in S}}, \tag{14}$$

where S denotes the same family of grid boxes as before (due to the previously mentioned spatial aggregation of the the population raster), and $count_p$ denotes the population count for grid box p.

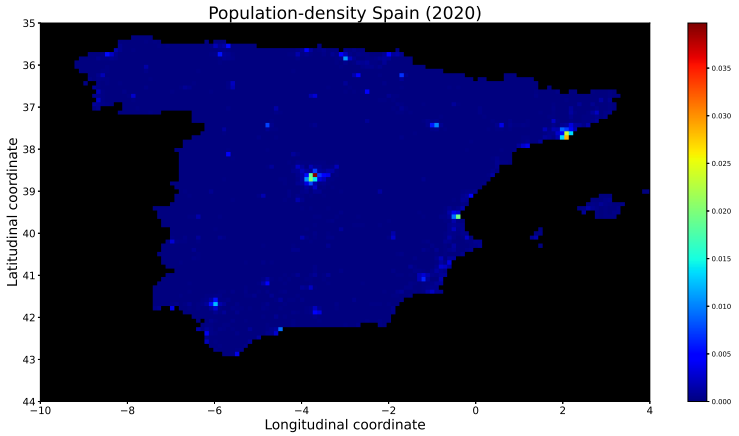


Figure 5: **Population-density data** (as percentage of total population) for Spain 2020.

As mentioned, population count data is available as a snapshot every 5 years, starting from year 1975 until year 2020. In the interest of time, the population data for year 2020 will serve as a proxy for the whole period from 2014 and onwards (as will be further discussed in future research points, a potential improvement of the methodology could be to use multiple population density rasters and linearly interpolate the data in-between these, taking into account more precisely the population density and how it affects the electricity demand on a country-level). After applying the country mask and the above transformation of the population count data, the final population density grid for Spain is demonstrated in Fig. 5. It should be noted that the electricity demand data of the islands on the east-coast of Spain are not considered in this research, however, due to the relatively small impact of the population count in these areas, these are not excluded in the final population density grid.

4.4 Local models

Given the disparity of Peninsular Spain in terms of geography and climate, as discussed in a previous section, it is reasonable to suggest that different degree-day variable thresholds should be applied to different parts of the country. As mentioned in Section 4.2, the objective was initially to utilise regional observed electricity demand data and train separate local models, whose combined output would be considered the estimate for the entire Peninsular Spain's electricity demand. Again, due to limited data supply, such an approach turned out to be infeasible. To this end, in addition to the construction of the novel split degree-day method, two approaches that could potentially circumvent this obstacle shall be considered. First, a naïve split of the country into four different regions will constitute the basis for a first localised version of the country model, where each region is optimised for independently using a locally optimal set of thresholds. Secondly, a more sophisticated clustering approach will be considered, utilising the k-means machine learning technique in order to split the country into different regions by the means of a similarity measure.

For both approaches, the training phase process is similar. It constitutes of tuning the threshold parameters for one region at a time by considering all possible combinations of CDD $= \{15, 17.5, 20, 22.5, 25\}$ and HDD $= \{10, 12.5, 15, 17.5, 20\}$, while keeping the other regions' parameters fixed to the thresholds deemed optimal for the corresponding country-level model. For the region being optimised for, the best set of threshold parameters is determined based on the R2-score measure-of-fit in the 10-fold CV. This procedure allows for locally optimised models without access to regional electricity demand data. It shall be noted that the R2-test score, as well as the MAPE 10-fold CV and test scores, are also computed and documented, but will not influence the final choice of tuned degree-day parameters.

4.4.1 Local models: arbitrary geographical split

As previously explained, the first local approach entails splitting the country into four arbitrary regions; North-East (NE), North-West (NW), South-East (SE), South-West (SW), with the aim to apply for each region optimal degree-day thresholds. The geographical division is arbitrarily carried out by selecting the middle coordinates in the longitudinal as well as the latitudinal direction of the country mask. As is visible in Fig. 6, such division will result in a disparity in the sizes attained by each region. For instance, the North-West region include many more nodal elements as compared to South-East region. To account for this, the degree-day variables will for each region be scaled by the size of the total number of nodal elements the region covers; that is, the degree-day variables of a large region will be assigned greater significance

than that of a small region when aggregating to country-level variables. Obviously, in order to enable a population-weighting of the degree-day variables of each region, the population raster must be split identically (see Fig. 14 in Appendix). The arbitrary split of the country into the four regions is visualized in Fig. 6.

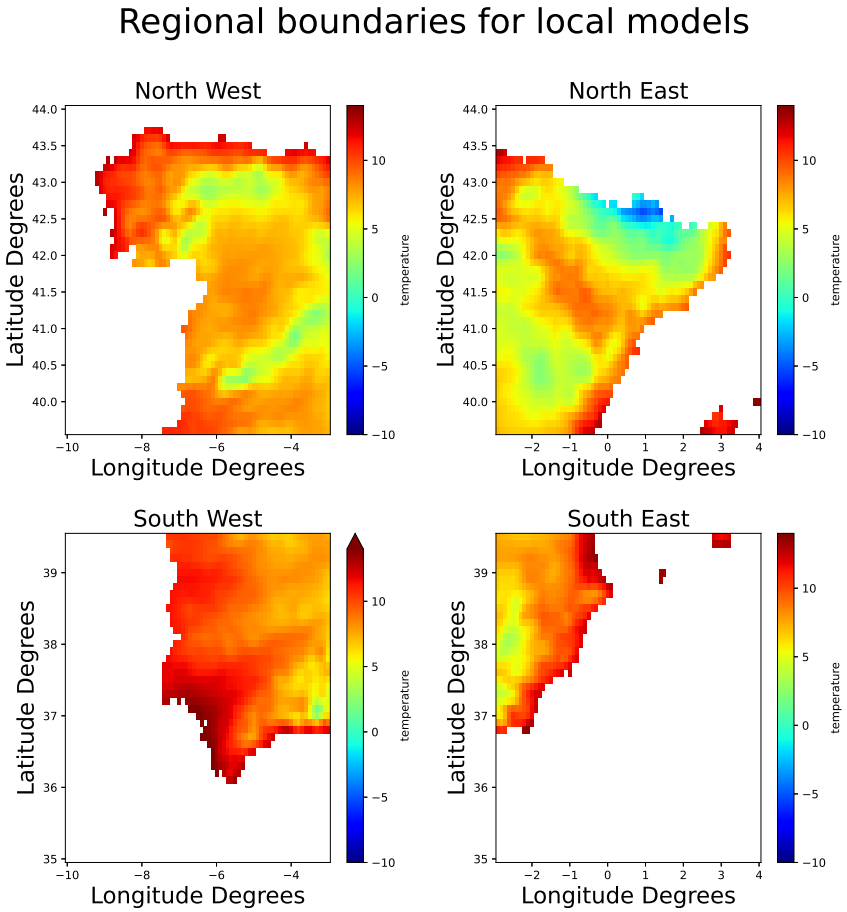


Figure 6: Snapshot (2014-01-01) of regional boundaries for four arbitrary regions of Peninsular Spain.

4.4.2 Local models: k-means clustering approach

Granted that the arbitrary division produced by the horizontal and vertical center points of the country mask achieves promising results, it would give reason to believe that a more thoughtful approach could yield superior results. To this end, as opposed to dividing the map of Peninsular Spain into four arbitrary regions, a more sophisticated approach to realize the split will follow hereafter, rooted in the k-means clustering approach.

The k-means clustering method, as previously described, works by considering the similarity between vectors in a d-dimensional vector space, given a predefined similarity measure. A common such similarity measure is the Euclidean distance between points in the space:

$$d(p, q) = \sqrt{\sum_{i=1}^{d} (q_i - p_i)^2}. \tag{15}$$

Given the assumption that the temperature in any given region is essential in determining the 'true" threshold values for the CDD and HDD indicators, it is a reasonable approach to use temperature as the key component in separating the country into different clusters. For this purpose, the calendar year shall be divided into what is known as the meteorological seasons:

- DJF: December to February,
- MAM: March to May,
- JJA: June to August,
- SON: September to November.

To this end, every grid point will be represented as a 4-dimensional vector, where each component contains the average temperature value for the specific season (calculated on hourly temperature data for the period 2014-01-01 - 2023-12-31). The number of desired clusters will be specified beforehand, and is usually denoted by k.

The k-means algorithm works by selecting centroids (centers of mass), thereafter measuring the distance (Euclidean, in this case) between the centroids (the number of centroids being equal to k - one for each cluster) and every point in the 4-dimensional space defined as in Eq. (15). Subsequently, each point in the space is assigned to the cluster corresponding to the centroid with which it is most similar to (i.e., the centroid with which the Euclidean distance between the point and the centroid is minimized). A new set of centroids is thereafter computed as the new center of mass of this updated cluster, and the iterations proceed until convergence (i.e., until the cluster assignments do not change from one iteration to another). For the first iteration, k random centroids are selected.

4.5 Model universe

The idea in this section is to present the model universe; in addition to the baseline model (the model whose results will serve as the benchmark for the results produced

by any other model presented herein), the model universe constitutes of all possible models that are constructed by modifying the baseline model by:

- the split degree-day approach (4.2), or
- a population-weighting approach (4.3, or
- the local arbitrary modelling approach (4.4.1), or
- the local k-means modelling approach (4.4.2), or
- a combination of two or more of the approaches above.

The model performance will be evaluated based on the two previously introduced well-known and frequently-used error measurements; R2-score and MAPE-score. The evaluation will be done in two stages. First, the model performance shall be evaluated using the 10-fold Cross-Validation approach seen previously, where 10 percent of the training data is repeatedly held out (the model is trained on the remaining 90 percent of the training data) as a validation set, which is then compared against corresponding predictions. As this is done 10 times, the resulting validation scores (for both R2 and MAPE) is computed as the average of scores of all folds. As adjacent points in a non-stationary time series are probable to be correlated with each other, the first and last two values in each validation set will be excluded from the validation score calculation. Secondly, the model is trained once more, but on the entire training set, and evaluated against the test data which it has not seen before. It shall also be emphasized that the division of the data, both the explanatory variables as well as the target variable, is made using the same parameterization in terms of history of the data, split-size and random state, enabling a fair comparison between all models.

4.6 Baseline model - BL

The Baseline model (BL) is the model as defined in Eq. (1), with threshold parameters of 15.5 and 22 for HDD and CDD variables, respectively. In order to get a country-level estimation of the electricity demand, the degree-day variables will be averaged across the entire grid.

4.7 Optimised Baseline model - OBL

This model is identical to the BL model, however, for the OBL model the threshold parameters are determined by considering the parameterization yielding the highest 10-fold CV R2-score.

4.8 Population-weighted baseline model - PBL

The Population-weighted Baseline model (PBL) is constructed by weighting the grid-level degree-day outputs by the corresponding population density. Thereafter, all grid point values across the grid are aggregated. The population density data has already been normalised to sum to 1.0 at this stage, thus no additional measures are necessary.

4.9 Optimised Population-weighted Baseline model - OPBL

As for OBL, the OPBL model is identically setup as the PBL, with the addition of using the optimal threshold parameters as opposed to the conventional thresholds.

4.10 Novel degree-day model - NDD

The Novel degree-day model is similar to the Baseline model, but with the difference that instead of using the conventional degree-day calculation, the novel split degree-day calculation as presented in Eqs. (7)-(10) will be applied.

4.11 Optimised Novel degree-day model - ONDD

The ONDD model corresponds to the NDD model, with the exception that instead of using the conventional thresholds of HDD = 15.5 and CDD = 22.0, the thresholds are found through the same optimization procedure as done for the OBL model.

4.12 Population-weighted Novel degree-day model - PNDD

This model is identical to the PBL model with the exception that the degree-day variables are calculated using the novel degree-day approach (as opposed to the conventional approach).

4.13 Optimized Population-weighted Novel degree-day model - PNDD

As for the ONDD model, the OPNDD model corresponds to the PNDD model with optimal threshold parameters.

4.14 Local conventional degree-day model - LCDD

The LCDD model is essentially the BL model divided into four subset; each subset is optimised for one by one as described in Section 4.4.1.

4.15 Local Novel degree-day model - LNDD

This model is identically setup as the LCDD model, except for the way in which the degree-day variables are calculated; for the LNDD, this is done using the novel degree-day approach. However, due to the intrinsic construction of the novel degree-day calculation scheme, hourly temperature data is required for the degree-day calculation of the LNDD model. Consequently, the computational load is more extensive, and thus, tuning the CDD and HDD parameters will require unreasonable time (considering roughly $90,000$ time steps, where each time step includes the computation of HDD and CDD variables of $\approx 10,000$ nodal coordinates). To this end, the same local threshold parameters as found for the LCDD model will be applied in this case as well. It is, however, with background to the intrinsic robustness of the novel degree-day approach that this decision is taken; indeed, the split-hour and window-size parameters to be derived will represent the configuration of the Novel degree-day model that is relatively insensitive to changes in threshold parameters, while achieving a relatively good measure-of-fit.

4.16 Local Population-weighted conventional degree-day model - LPCDD

The LPCDD model is the result of applying a population-weighting to the degree-day values at each grid point calculated using the conventional degree-day approach, before optimising for the regions one at a time.

4.17 Local Population-weighted Novel degree-day model - LPNDD

This setup is identical to that of the LPCDD model, with the difference lying in the method with which the degree-day variables are calculated; LPNDD model uses degree-days calculated using the novel split degree-day approach.

4.18 Local k-means conventional degree-day model - LKCDD

The LKCDD model is similar to the LCDD model, with the additional feature that clusters are derived using the previously introduced k-means clustering approach on temperature data. For each k, the regions will be optimized for as in the LCDD model; by updating the HDD and CDD threshold parameters based on the best R2 measure-of-fit for one cluster at a time, keeping all other clusters' threshold parameters fixed to the optimal threshold as found for the OBL model. After the optimal thresholds for a cluster have been found, the HDD and CDD parameters are updated to these thresholds before the next cluster is considered. Due to the randomness in the model

stemming from which initialisation points are selected, each k will be experimented for twice; each time with different initial centroids, and the final R2 and MAPE scores will be computed as the average of those achieved for each of the two experiments. Furthermore, for each experiment, the order in which the clusters are being optimized for will depend on the initial centroids, and might therefore be scrambled. As previously mentioned, the contribution of each cluster will be weighted by the number of nodal coordinates included in the cluster. The clustering will be done for $k \in \{2, 3, 4\}$.

4.19 Local Population-weighted k-means conventional degree-day model - LPKCDD

The LPKCDD model is constructed similarly to the LKCDD model, however, instead of weighting the contribution of each cluster by the corresponding cluster size, the degree-day variables calculated for each nodal coordinate will be weighted by the corresponding population density for that same coordinate.

5 Experimental evaluation

When evaluating the results, a train and validation period has been determined to be between 2014-12-19 - 2018-05-03, whereas the test period is between 2018-05-04 - 2019-12-31. This split between train and test data is consistent for all results provided herein.

5.1 Split degree-day method - optimal configuration

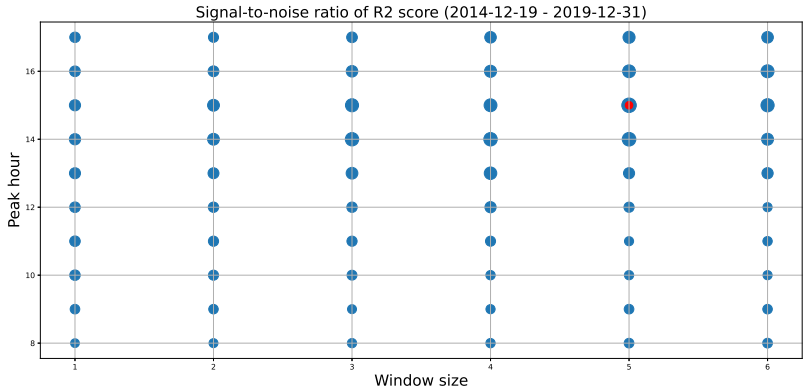


Figure 7: Each point in the plot is calculated as the median of all possible combinations of degree-day thresholds for CDD and HDD variables over the interquartile range ($IQR = Q_{75} - Q_{25}$). The radius of the circle represents signal strength = $\frac{median}{IQR}$, and the highest value is marked by the red dot.

Firstly, we shall present the results associated with the configuration of the split degree-day method. In terms of maximizing the signal strength given by Eq. (13), the optimal configuration, as visualised in Fig. 7, proved to be peak-hour := 15:00 and window-size := 5. Considering Fig. 11 in Appendix, it is obvious that this configuration yields a peak interval which captures both of the electricity load peaks for the average 24-hour period, whereas, if considering Fig. 10 in Appendix, the peak-hour of 15:00 is precisely the peak-hour of the average temperature curve of any given day. This result is aligned with the theory of the split degree-day method as presented earlier. As a side note, it should be mentioned that this approach of calculating the signal strength assigns equal importance to all scores, indifferent to what combination of the threshold parameters being used; even combinations that could be deemed "less likely", for instance, the pair {HDD,CDD} = $\{20, 10\}$, are assigned the same importance as more probable threshold combinations. Therefore, it was also experimented with assigning the weights to each score (one score is obtained for each combination of thresholds) as derived by a Gaussian distribution, assigning greater significance to

elements located more centrally in the sets CDD = $\{15.0, 17.5, 20.0, 22.5, 25.0\}$, HDD = $\{10.0, 12.5, 15.0, 17.5, 20.0\}$. However, exactly the same peak-hour and window-size configuration was obtained with this alternate method as well.

5.2 Parameterization of local models

5.2.1 Local arbitrary split

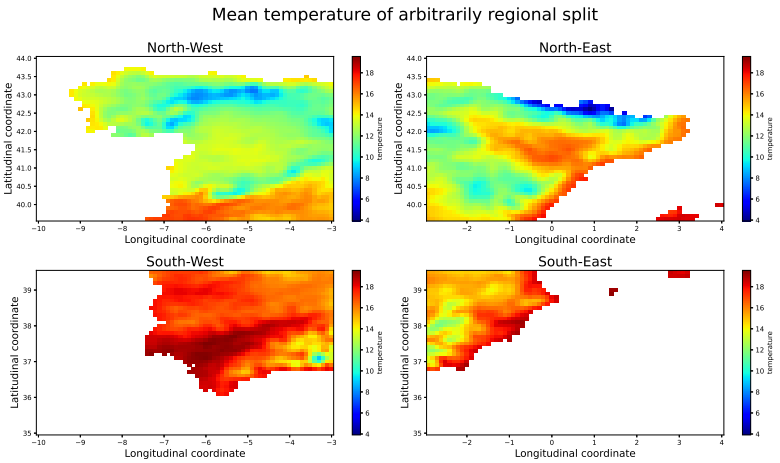


Figure 8: Mean temperatures of Spain, arbitrary regions highlighted.

As becomes obvious by inspecting Fig. 8 and Tables 3-4, cooler areas (North-West and North-East) are associated with lower optimal HDD thresholds and higher optimal CDD thresholds, suggesting that such areas are more inclined to applying heating at cooler temperatures, whereas cooling is first applied for relatively high temperatures. For the warmer areas (South-West and South-East), population-weighting is apparently significant in determining the optimal thresholds, resulting in relatively similar base-temperatures for both heating and cooling.

Table 3: Parameterization - no population-weighting

Region	HDD	CDD
North-West	12.5	25.0
North-East	12.5	22.5
South-West	20.0	17.5
South-East	17.5	15.0

Table 4: Parameterization - with population-weighting

Region	HDD	CDD
North-West	15.0	25.0
North-East	12.5	22.5
South-West	17.5	20.0
South-East	17.5	17.5

5.2.2 k-means clustering split

Secondly, the resulting clusters generated by the k-means algorithm (the maps generated by each seasonal temperature component are visualised separately for each season

in Appendix 15-18) for $k \in \{2, 3, 4\}$ look, for the an arbitrary random state, as follows:

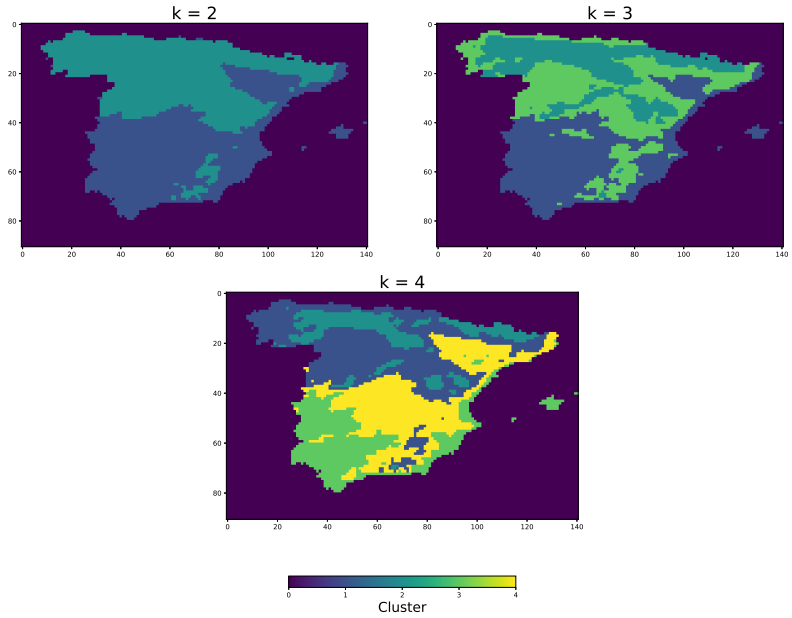


Figure 9: Clustering achieved for $k \in \{2, 3, 4\}$ based on mean seasonal temperatures. Cluster 0 is merely included to facilitate computations (representing water) and can be disregarded.

The paramerization corresponding to each cluster is given by tables 5-7 (with p.w. indicating population-weighting has been applied):

Table 5: Parameterization - k=2

Cluster	HDD	CDD	HDD-p.w.	CDD-p.w.
1	10.0	22.5	12.5	22.5
2	17.5	17.5	15.0	20.0

For the parameterization when $k = 2$ in Table 5, if considering the mean temperature as depicted in Fig. 8 and the clusters presented in Fig. 9, it is clear that, without population-weighting, the generally cooler temperatures of cluster 2 yield a lower HDD threshold and a higher CDD threshold than that of cluster 1. However, as the degree-days are weighted by the population density, the difference between the optimal thresholds is less extensive, but the same pattern is still present.

For $k = 3$, with corresponding parameterization visible in Table 6, cluster 2, the clus-

Table 6: Parameterization - k=3

Cluster	HDD	CDD	HDD-p.w.	CDD-p.w.
1	20.0	17.5	20.0	17.5
2	12.5	22.5	20.0	22.5
3	15.0	25.0	12.5	25.0

Table 7: Parameterization - k=4

Cluster	HDD	CDD	HDD-p.w.	CDD-p.w.
1	12.5	22.5	12.5	25.0
2	20.0	22.5	10.0	25.0
3	20.0	17.5	20.0	17.5
4	17.5	15.0	15.0	25.0

ter with an on average cooler temperature, displays a similar behaviour as cluster 2 for $k = 2$; optimal HDD threshold is lower than those of the 2 warmer clusters, with optimal CDD being significantly higher, and the difference between the two shrink as population-weighting is applied. The on average warmest cluster, $k = 1$, displays the opposite pattern; both with and without population-weighting, optimal HDD threshold is higher than optimal CDD threshold, however, the difference between the two is not as significant. Probably, this indicates that the dependence on the electricity demand of heating is less significant in warmer areas, whereas cooling is generally applied at lower temperatures. Lastly, for $k = 3$, the cluster for which temperature is warm/cool, the same pattern as the one prevalent for the coolest cluster is distinguishable, with all thresholds, with and without population weighting, slightly higher.

For the clustering when $k = 4$ with parameterization in Table 7, the coolest cluster, 2, is again displaying the same pattern as before, however, the application of population-weighting is seemingly more significant in this case, pushing optimal HDD down and optimal CDD up. The on average warmest cluster, 3, again shows a low threshold for CDD and a relatively high threshold for HDD. The remaining clusters, 1 and 4, show similar optimal thresholds as population-weighting is applied, whereas the relationship between the thresholds are otherwise flipped. Perhaps this indicates that densely populated areas (such as Madrid) in cluster 4 are more prevalent to apply heating at lower temperatures, whereas the effect of cooling is less significant.

5.3 Model performances

Further, a comprehensive view should be given of the results of all models, in terms of their 10-fold CV scores as well as the out-of-sample (OOS) test scores. This is given below, with the top performers for both the R2 and the MAPE scores, as well as for both training and test categories, highlighted in bold-face. Before reviewing the results of this section more thorough, it serves to clarify the distinction between the 10-fold CV (train) and OOS (test) categories. For both categories, the models' performances are calculated based on predicting the electricity demand using input data they have not been trained on. However, for each of the 10 folds in the CV, the held-out sample is considerably smaller than the OOS. Also, by the manner in which the CV is defined, for each held-out sample, except for the first and last one, the model is trained on data on both sides (before and after) of the sample it is tested on, which is not the case for the OSS, which includes data with start date being located after the end date of the training data. In order to make the review comprehensive yet reasonable in length, an in-depth review shall be limited to some specific categories.

MODEL	Acronym	**10-fold CV**	**OOS**
Baseline model	BL 4.6 Bloomfield et al. [2021]	0.7263	0.6666
Optimised Baseline model	OBL 4.7	0.7403	0.6993
Pop-weighted Baseline model	PBL 4.8	0.7405	0.6751
Optimised Pop-weighted Baseline model	OPBL 4.9	0.7453	0.6852
Novel degree-day	NDD 4.10	0.7327	0.6809
Optimised Novel degree-day	ONDD 4.11	0.7369	0.7010
Pop-weighted Novel degree-day	PNDD 4.12	0.7443	0.6852
Optimised Pop-weighted novel degree-day	OPNDD 4.13	0.7444	0.6848
Local conv. degree-day	LCDD 4.14	0.7438	0.6977
Local Pop-weighted conv. degree-day	LPCDD 4.16	**0.7512**	0.6869
Local Novel degree-day	LNDD 4.15	0.7408	**0.7034**
Loc. Pop-weighted Novel degree-day	LPNDD 4.17	0.7474	0.6917
Loc. k-means conv. degree-day (k=2)	LK2CDD 4.18	0.7454	0.7015
Loc. Pop-w. k-means conv. degree-day (k=2)	LPK2CDD 4.19	0.7478	0.6866
Loc. k-means conv. degree-day (k=3)	LK3CDD 4.18	0.7461	0.6981
Loc. Pop-w. k conv. degree-day (k=3)	LPK3CDD 4.19	0.7480	0.6856
Loc. k-means conv. degree-day (k=4)	LK4CDD 4.18	0.7456	0.6991
Loc. Pop-w. k-means conv. degree-day (k=4)	LPK4CDD 4.19	0.7498	0.6846

Table 8: Score table - R2.

MODEL	Acronym	**10-fold CV**	**OOS**
Baseline model	BL 4.6 Bloomfield et al. [2021]	0.03515	0.03666
Optimised Baseline model	OBL 4.7	0.03274	0.03413
Pop-weighted Baseline model	PBL 4.8	0.03342	0.03469
Optimised Pop-weighted Baseline model	OPBL 4.9	0.03255	0.03343
Novel degree-day	NDD 4.10	0.03411	0.03571
Optimised Novel degree-day	ONDD 4.11	0.03329	0.03414
Pop-weighted Novel degree-day	PNDD 4.12	0.03275	0.03408
Optimised Pop-weighted Novel degree-day	OPNDD 4.13	0.03250	0.03361
Local conv. degree-day	LCDD 4.14	0.03245	0.03426
Local Pop-weighted conv. degree-day	LPCDD 4.16	0.03165	0.03376
Local Novel degree-day	LNDD 4.15	0.03275	0.03422
Local Pop-weighted Novel degree-day	LPNDD 4.17	0.03207	0.03406
Loc. k-means conv. degree-day (k=2)	LK2CDD 4.18	0.03239	0.03366
Loc. Pop-w. k-means conv. degree-day (k=2)	LPK2CDD 4.19	0.03182	0.03366
Loc. k-means conv. degree-day (k=3)	LK3CDD 4.18	0.03226	0.03364
Loc. Pop-w. k-means conv. degree-day (k=3)	LPK3CDD 4.19	0.03184	0.03383
Loc. k-means conv. degree-day (k=4)	LK4CDD 4.18	0.03219	**0.03337**
Loc. Pop-w. k-means conv. degree-day (k=4)	LPK4CDD 4.19	**0.03157**	0.03380

Table 9: Score table - MAPE.

5.3.1 Not optimized models

By inspecting Tables 8-9, some facts become apparent. Firstly, in terms of both score measures, it can be concluded that the worst performing model is in fact the Baseline model (BL), ranking last out of all models for both the train and test sample. However, as is emphasized in Section 4.6, this model, along with NDD, LNDD, LPNDD (see Sections 4.10, 4.15, 4.17), are the only models not being optimised for. Comparing among them, the BL model will still rank last, followed by the NDD model. For the R2-score, however, the LNDD model achieves the highest test score of all models. This could probably be attributed to the fact that the split degree-day method, the intrinsic degree-day method utilised by this model, has been configured with the objective of being relatively insensitive to the inputted baseline thresholds, while still achieving a relatively high measure-of-fit. Hence, it stands to reason that the LNDD model generalises relatively better to unseen test data, thus achieving the best OOS score.

5.3.2 Best performing models - 10-fold CV

In terms of the best performing models on the 10-fold CV set, localized versions are generally superior. Probably, this owes to the increased flexibility in such models, due to the fact that each region is optimised for separately. For the MAPE score, the LPK4CDD model is superior to all other models. It is also clear that, as could be expected, the 10-fold CV score improves with the number of clusters, which again

could be a cause of the increased flexibility of the model. For the R2 measure-of-fit, the best performer on the 10-fold CV is the LPCDD model. Considering the arbitrary split of the country into local regions for this model, this result is somewhat surprising, as it suggests a better measure-of-fit is achieved by splitting the country in this manner, as opposed to clustering the country into regions by considering the general temperature pattern. Further, it should be noted that weighting the degree-day outputs with the corresponding population density, for all models and both score measures, increases the performance in the CV.

5.3.3 Best and worst performing models - test

In regards to the test data, the distribution of the performance of models is apparently slightly different from what was presented for the CV scores. As previously mentioned, in terms of the R2-score, the LNDD model ranks on top. Also the LK2CDD and ONDD models show notable results. In terms of the LK2CDD model, the performance could be attributed to the fact that a clustering of $k = 2$ counteracts potential overfitting to the data as seen higher up in the cluster hierarchy. For the ONDD, this could again be a product of the way the split degree-day method was configurated. On the other hand, the worst performing models OOS, except for the Baseline model, is the Optimised Baseline model, followed closely by the NDD model. For the latter of the two, it could potentially be a case of underfitting, given that the model has not been optimised for in terms of threshold parameters.

5.3.4 General results

In general, for both score types and sample categories, it is clear that the differences between the models' performances are not that significant. Perhaps, given that the construction of each model is rather different from any other, this is an indication of the shortcomings of the underlying multiple linear regression model that is being applied as the predictor mechanism for each model in this universe. It also suggests that the non-weather-dependent coefficients of the regression model (i.e. all variables except the degree-day variables) are imperative in predicting the electricity demand, leaving a relatively lower significance to the degree-day variables as the models are fitted. Thus, even if the degree-day variables for all models are intrinsically different in their construction, their impact on the final result is marginalized by the relatively large contribution of the variables the models have in common (again, all variables except those of degree-days). An experiment using de-seasonalised electricity demand, predicted solely by weather-dependent variables (i.e. the degree-day variables) as well as an intercept, is briefly discussed in Section 8.2.1 in Appendix.

By inspecting Figs. 19 20 in Section 8.2, it is clear that all models produce degree-day

variables of relatively high correlation when compared to each other; all correlations being higher than 0.9, which is undoubtedly another reason for the similar performances of the different models.

It shall also be mentioned that when fitting the multiple linear regression model without intercept, the R2-score can appear inflated due to the increased SS_{tot} term in Eq. (5) as $\hat{y}$ is removed, however, leading to substantially worse predictions. Importantly, all regression models in this research are fitted with intercept.

6 Conclusion

This dissertation has attempted to provide an extensive assessment of the potential impact of applying promising research findings and novel concepts to the skillful electricity demand prediction model as presented by [Bloomfield et al., 2021]. Particularly, this research has been centred around the use of the so-called degree-day variables in predicting country-level electricity demand for Spain, and exploring the ideas of the novel split degree-day method presented in [Kheiri et al., 2023], the approach of weighting degree-days by the corresponding population density of the specific area, as well as attempting to increase the flexibility of the country-level prediction model by exploring with clustering approaches to identify similar regions within the country and taking into account the heterogeneous climate across these.

This study has shown that improvements to the Baseline model 1 are possible. Specifically, the application of optimised degree-day variables in terms of optimal threshold parameters has proven to be effective. Moreover, it has been shown that there is an apparent benefit in including the element of population-weighting when predicting electricity demand as a function of temperature. Even if the clustering, based on temperature data, of the country into a set of distinct regions has shown to be a capable approach, the added value is apparently less significant than what was expected. In general, given that the degree-day input variables to the Baseline model are optimised and population-weighted, the added effect of extending the method to include locally optimised clusters, regardless in which way these clusters are attained by (i.e., whether the clustering is arbitrary or by applying the k-means clustering approach), is generally negligible.

7 Future work

There are areas related to this research that can be further explored. For instance, as is depicted by Figs. 19- 20 in Appendix, the resulting degree-day variables, across all models, are relatively similar. However, given that the methods by which they have been computed are rather different, including more countries in the research would be interesting. Further, another promising research area that could be pursued is to consider a more flexible prediction mechanism than the multiple linear regression model given in Eq. (1). Perhaps, increased flexibility of the model could lead to enhanced accuracy. In order to increase reliability of the population-weighting approach, the use of multiple population count rasters might also prove useful. Lastly, given that there is reason to believe optimal thresholds for the degree-day variables are not only dependent on the spatial location for which the variables are calculated, but also of the time of the year, considering time-varying thresholds could be fruitful in further

optimizing the prediction model.

8 Appendix

8.1 Data

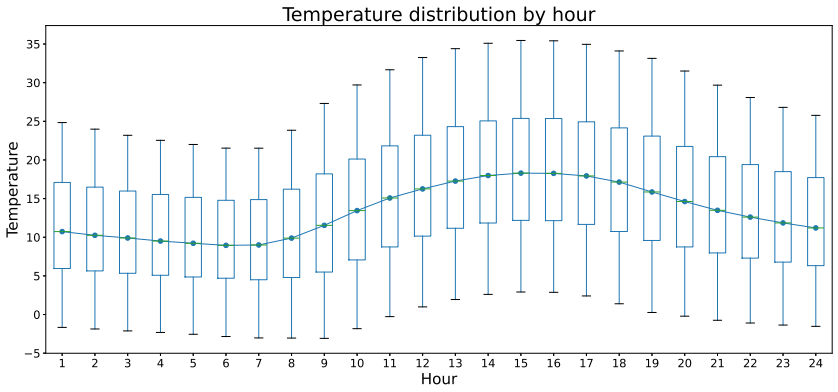


Figure 10: Temperature distribution by hour (calculated on data 2014-12-19 - 2019-12-31). Code from: [Amat-Rodrigo and Escobar-Ortiz, 2024].

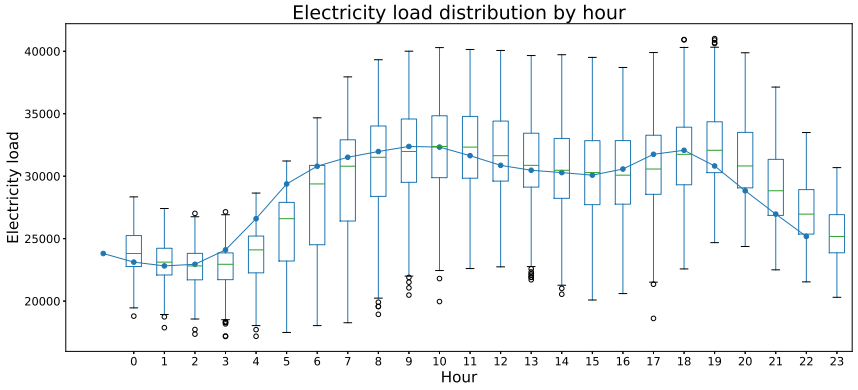


Figure 11: Electricity load distribution by hour (calculated on data 2014-12-19 - 2019-12-31). Code from: [Amat-Rodrigo and Escobar-Ortiz, 2024].

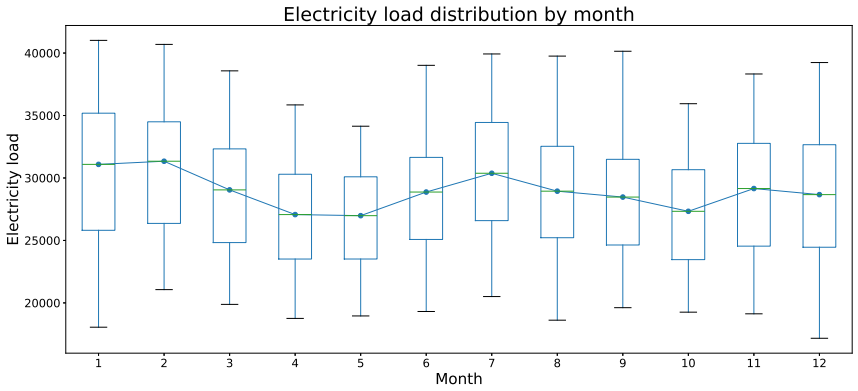


Figure 12: Electricity load distribution by month (calculated on data 2014-12-19 - 2019-12-31). Code from: [Amat-Rodrigo and Escobar-Ortiz, 2024].

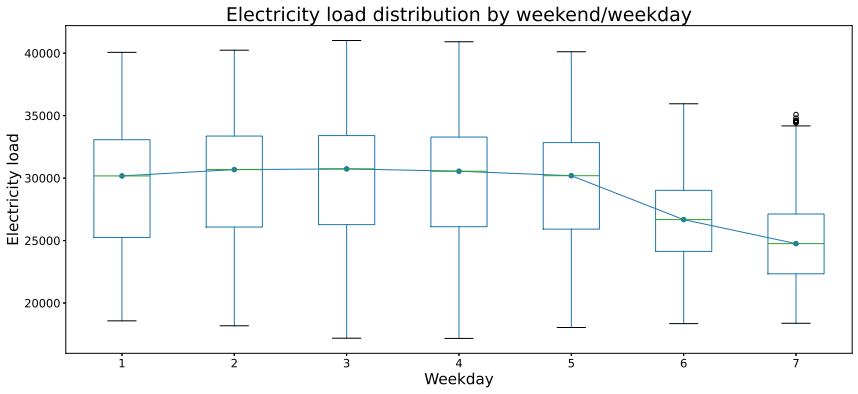


Figure 13: Electricity load distribution by weekend and weekday (calculated on data 2014-12-19 - 2019-12-31). Code from: [Amat-Rodrigo and Escobar-Ortiz, 2024].

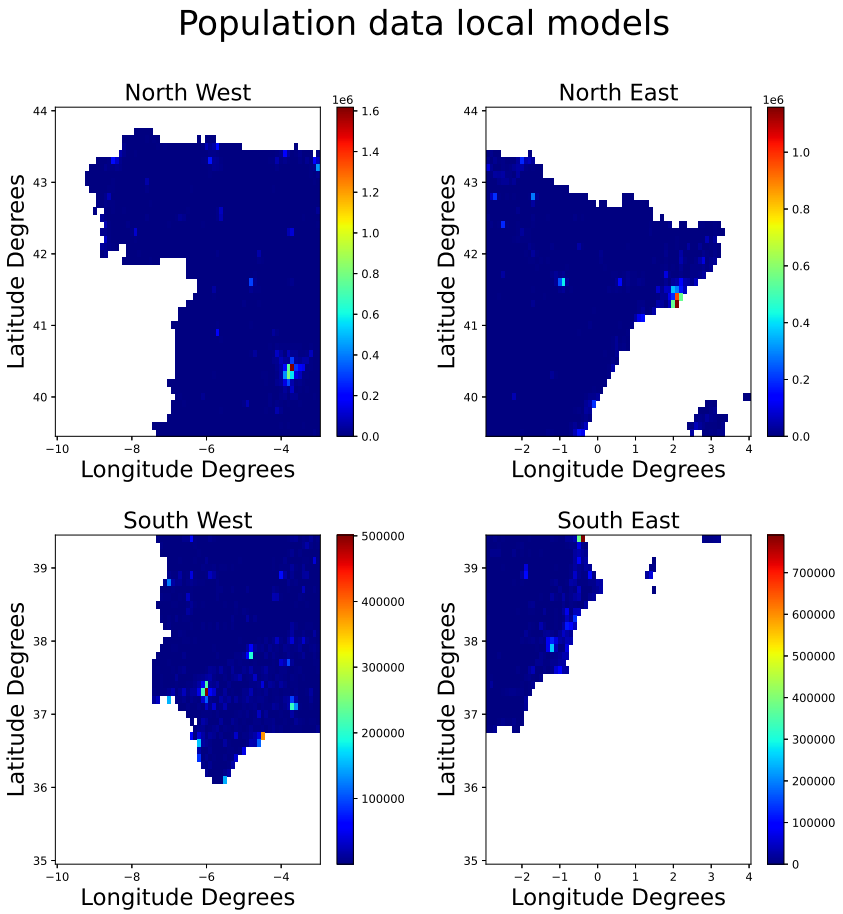


Figure 14: Population data (2020) for arbitrarily-split regional models for Spain.

8.2 Results

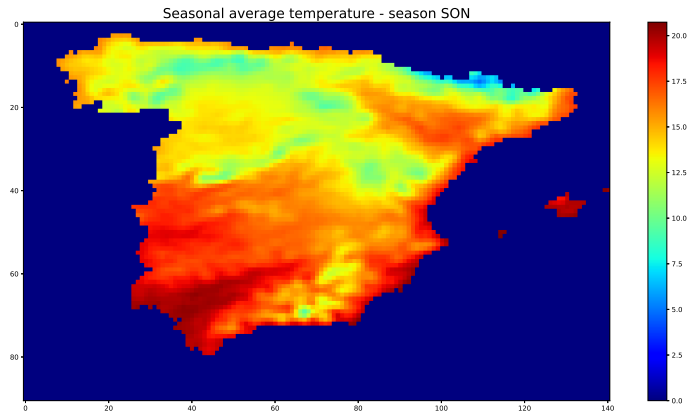


Figure 15: Mean seasonal temperature for Spain - season SON.

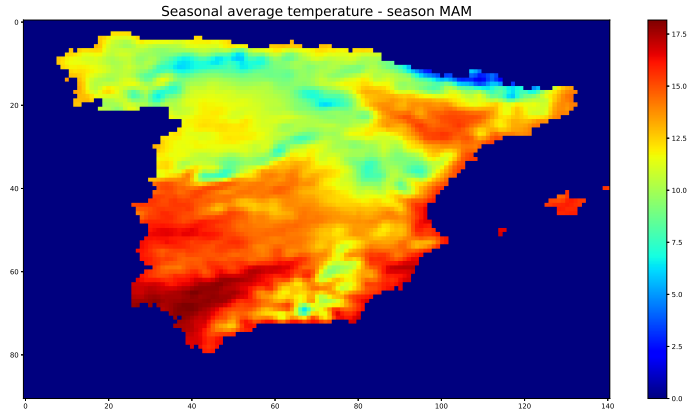


Figure 16: Mean seasonal temperature for Spain - season MAM.

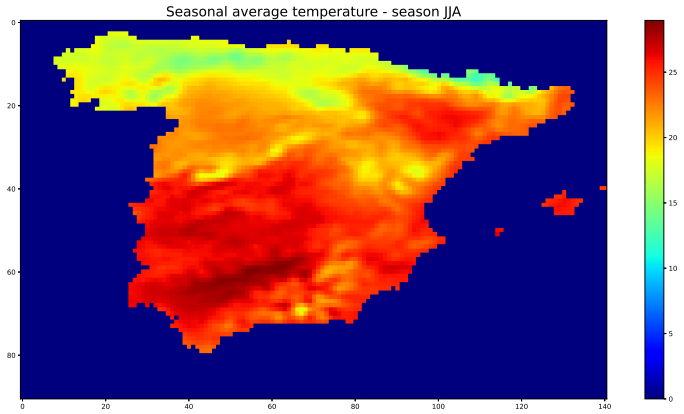


Figure 17: Mean seasonal temperature for Spain - season JJA.

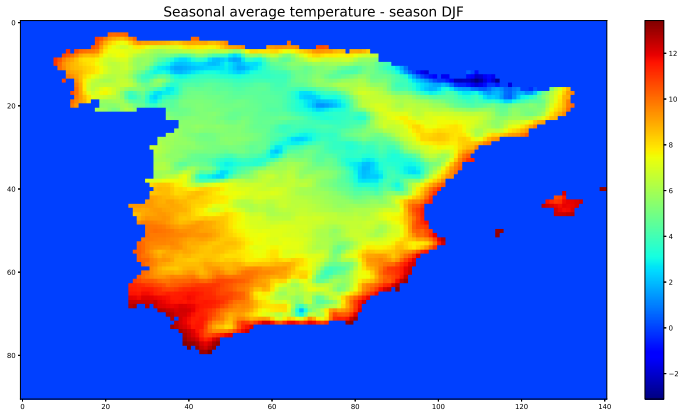


Figure 18: Mean seasonal temperature for Spain - season DJF.

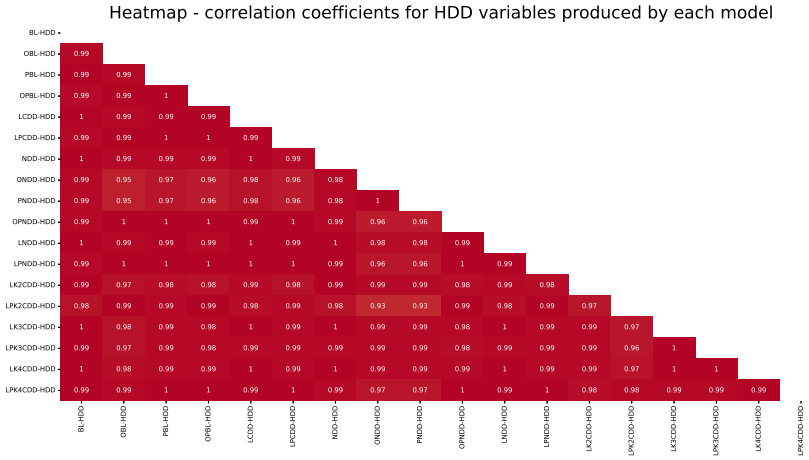


Figure 19: Heatmap of correlation coefficients calculated on HDD variables produced by each model.

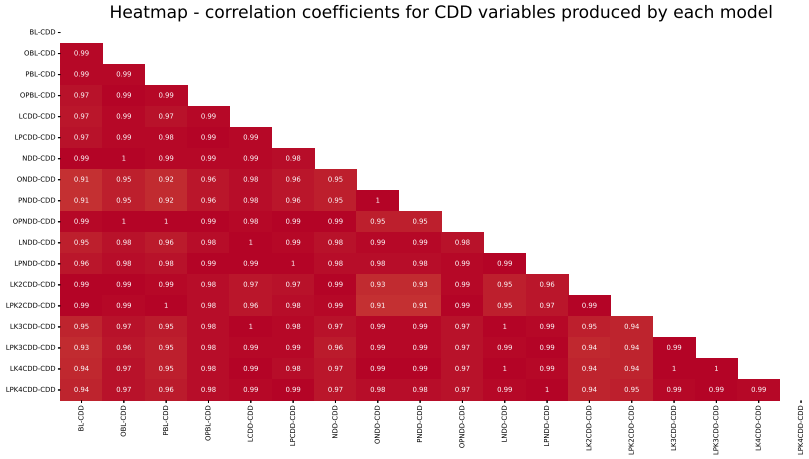


Figure 20: Heatmap of correlation coefficients calculated on CDD variables produced by each model.

8.2.1 Results for models fitted on de-seasonalised electricity demand

In order to combat the mentioned marginalization due to the weekly cyclical variation in electricity demand captured by the dummy variables in the regression model, results shall be presented where the dependent variable has been de-seasonalized (i.e. weekly variation removed) and predicted solely using the degree-day variables (in addition to the model intercept coefficient). The resulting electricity demand series has only a trend as well as a residual component, and in-sample train scores as well as out-of-sample test scores are visible for each model in Figs. 21-22. As is apparent, although differences amongst model are distinguishable to a larger extent in this way,

their performances are still relatively similar.

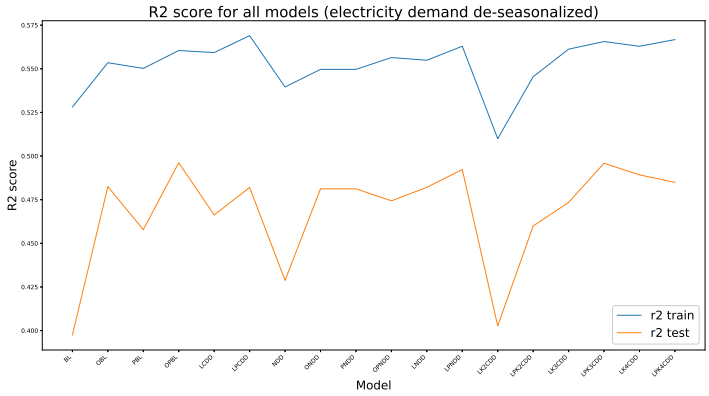


Figure 21

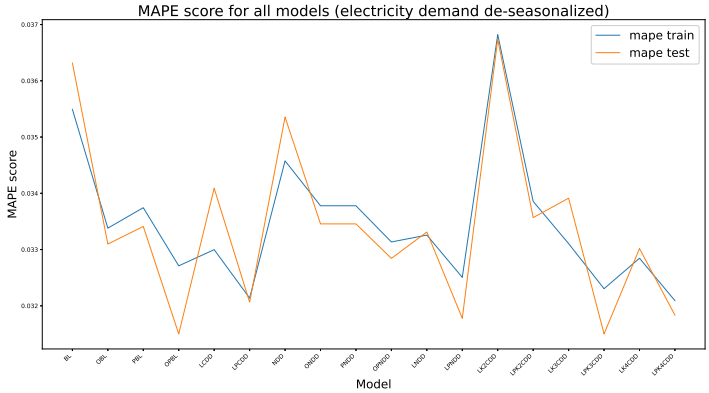


Figure 22

References

Joaquin Amat-Rodrigo and Javier Escobar-Ortiz. skforecast, 2 2024. URL `https://skforecast.org/`.

Juliana Antunes-Azevedo, Lee Chapman, and Catherine L. Muller. Critique and suggested modifications of the degree days methodology to enable long-term electricity consumption assessments: a case study in birmingham, uk. *Royal Meteorological Society*, 2015. doi: https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/met.1525.

Hannah C. Bloomfield, David J. Brayshaw, and Andrew Charlton-Perez. Characterizing the winter meteorological drivers of the european electricity system using targeted circulation types. 2020. doi: https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.1858.

Hannah C. Bloomfield, David J. Brayshaw, Paula L. M. Gonzalez, and Andrew Charlton-Perez. Sub-seasonal forecasts of demand and wind power and solar power generation for 28 european countries. 2021. doi: https://essd.copernicus.org/articles/13/2259/2021/.

Eduardo Caro, Jesús Juan, and Javier F. Cara. Periodically correlated models for short-term electricity load forecasting. *Applied Mathematics and Computation 00*, 2019. doi: https://oa.upm.es/79586/1/applied%20Math_EC_JJ_JC.pdf.

William Chung. Using the fuzzy linear regression method to benchmark the energy efficiency of commercial buildings. *Applied Energy*, 2012. doi: https://www.sciencedirect.com/science/article/pii/S0306261912000670?via%3Dihub.

European Commission. An eu strategy on heating and cooling. *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS*, 2016. doi: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016DC0051.

Copernicus. Heating and cooling degree days from 1979 to 2100. *Copernicus Climate Data Store*, 2022. doi: https://cds.climate.copernicus.eu/cdsapp#!/software/app-heating-cooling-degree-days?tab=overview.

Copernicus. Era5-land hourly data from 1950 to present. *Copernicus Climate Data Store*, 2024a. doi: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview.

Copernicus. Ghs-pop - r2023a. *GHSL - Global Human Settlement Layer*, 2024b. doi: https://human-settlement.emergency.copernicus.eu/ghs_pop2023.php.

EIA. Units and calculators explained. *U.S. Energy Information Administra-*

tion, 2023. doi: https://www.eia.gov/energyexplained/units-and-calculators/degree-days.php.

Bill Gates. It is surpringsingly hard to store energy. *Gates Notes*, 2016.

Tao Hong, Pierre Pinson, Yi Wang, Rafał Weron, Dazhi Yang, and Hamidreza Zareipour. Energy forecasting: A review and outlook. *Journal of Power and Energy*, 2020. doi: https://www.researchgate.net/publication/344890706_Energy_Forecasting_A_Review_and_Outlook.

IEA. Renewables 2023: The global power mix will be transformed by 2028. Technical report, International Energy Agency, 2O23.

Harry R. Kennard, Oreszczyn Tadj, Malcolm N. Mistry, and Ian G. Hamilton. Population-weighted degree-days: The global shift between heating and cooling. *Energy and Buildings*, 2022. doi: https://www.sciencedirect.com/science/article/pii/S0378778822004868#:~:text=Globally%2C%20population%20weighting%20substantially%20changes,cooling%20degree%20days%20are%20increasing.

Farshad Kheiri, Jeff S. Haberl, and Juan-Carlos Baltazar. Split-degree day method: A novel degree day method for improving building energy performance estimation. *Energy and Buildings*, 2023. doi: https://www.sciencedirect.com/science/article/pii/S0378778823002645.

Robert G. Quayle and Henry F. Diaz. Heating degree day data applied to residential heating energy consumption. *Journal of Applied Meteorology and Climatology*, 1980. doi: https://journals.ametsoc.org/view/journals/apme/19/3/1520-0450_1980_019_0241_hdddat_2_0_co_2.xml?tab_body=pdf.

Ying Shi, Xuejie Gao, Ying Xu, Filippo Giorgi, and Deliang Chen. Effects of climate change on heating and cooling degree days and potential energy demand in the household sector of china. *Climate Research*, 2016. doi: https://www.int-res.com/abstracts/cr/v67/n2/p135-149/.

Jonathan Spinoni, Jürgen V. Vogt, Paulo Barbosa, Alessandro Dosio, Niall McCormick, Andrea Bigano, and Hans-Martin Füssel. Changes of heating and cooling degree-days in europe from 1981 to 2100. *Royal Meteorological Society*, 2017. doi: https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/joc.5362.

Billie L. Taylor. Population-weighted heating degree-days for canada. *Taylor Francis*, 1981. doi: https://www.tandfonline.com/doi/abs/10.1080/07055900.1981.9649113.

H. C. S. Thom. The rational relationship between heating degree days and temperature. *American Meteorological Society*, 1954. doi: https://doi.org/10.1175/1520-0493(1954)082%3C0001:TRRBHD%3E2.0.CO;2.

Enric Valor, Vicente Meneu, and Vicente Caselles. Daily air temperature and electricity load in spain. *American Meteorological Society*, 2001. doi: https://journals.ametsoc.org/view/journals/apme/40/8/1520-0450_2001_040_1413_datael_2.0.co_2.xml.