Weather Effects in Energy Seasonal Adjustment: An Application to France Energy Consumption*

Marie Bruguet[†,1,2,3], Ronan Le Saout[‡4], and Arthur Thomas [§,1,3]

[1]Université Paris-Dauphine, Université PSL, LEDa, CNRS, IRD, 75016 Paris, France
[2]Commissariat Général au Développement Durable, SDES, 92700, Puteaux, France
[3]Climate Economics Chair, 75001, Paris, France
[4]Université de Rennes, ENSAI, CREST, F-35000 Rennes, France

September 5, 2024

Abstract

This paper addresses the challenge of accurately adjusting energy consumption data for weather and seasonal variations. In this paper, we revisit the concept of Heating Degree Days (HDD) by proposing a new General Weather Indicator (GWI). This indicator is a vector of the optimal combination of heating and cooling day variables across multiple weather variables, including temperature, wind, sunlight duration, rain, and cloudiness. To construct the GWI, we propose a novel econometric approach which uses K-means to identify thresholds and LASSO penalisation for the selection of variables. An empirical analysis of electricity and natural gas consumption in France reveals that there are nuanced sectoral responses to weather, highlighting the importance of integrating wind and sunlight duration along with temperature. The reliability of the GWI was confirmed by robustness checks across spatial and temporal scales, which demonstrated its suitability for use in different geographical areas and at different data frequencies. Using the traditional base temperature of 17°C instead of the optimal GWI for energy adjustment, we found an underestimated consumption response to temperature of up to 7.3%, or 0.80 TWh per month, equivalent to more than one and a half nuclear reactors, highlighting the importance of accurate weather elasticity estimates.

JEL classification: C32, E61, P28, Q47
Keywords: sectoral energy consumption, seasonal adjustment, weather, k-means clustering, LASSO penalisation, heating degree days

*We thank Virginie Andrieux, Marie Bessec, Anna Creti, Tomás Del Barrio Castro and Daniel Herrera Araujuo for many useful suggestions and detailed comments, as well as seminar and conference participants at the International Association for Applied Econometrics (IAAE 2024), 2024 RCEA International Conference in Economics, Econometrics, and Finance (ICEEF 2024) and 44th International Symposium on Forecasting (ISF) for their insights.

[†]email: marie.bruguet@dauphine.eu
[‡]email: ronan.le-saout@ensai.fr
[§]email: arthur.thomas@dauphine.psl.eu

1 Introduction

To mitigate the effects of climate change, the Paris Agreement calls for a reduction in global emissions. This highlights the significant potential for European countries to meet their emissions targets by significantly reducing energy consumption. In 2021, energy use accounts for 76.7% of total emissions on average, underscoring its key role in achieving environmental goals.[1] In this context, the French government has implemented a "Pluriannual Energy Plan" with the objective of reducing energy consumption by 20% in 2030 compared to 2012 (see MTE, 2020). This initiative aims to reduce both greenhouse gas emissions and dependence on fossil fuels. In order to assess the effectiveness of such strategies, government agencies need to use tools that facilitate the evaluation of their impact in real time. Although it is possible to evaluate the causal effect of a specific efficiency policy aimed at reducing energy consumption in a posteriori manner, this is not a straightforward process, (see e.g. Gerarden et al., 2017; Fowlie et al., 2018; Peñasco and Anadón, 2023). The assessment of changes in energy consumption in real time represents a significant challenge, given the multitude of factors that contribute to it, including socio-economic, political, and climatic factors (see e.g. Dell et al. (2014); Aubin et al. (1995)).

This paper focuses on identifying and analysing the various factors that influence energy consumption, with particular emphasis on seasonal and meteorological variables. It is evident that there is a significant correlation between monthly energy consumption and temperature. This is largely due to the high energy demand for heating buildings during the winter months and cooling buildings in hot climates during the summer. To ensure reliable comparisons over time and to assess the impact of socio-economic factors alone, it is essential to correct observed consumption for seasonality and climatic variations (see e.g. Pang et al., 2022). It is therefore important for public policy makers to have access to robust seasonal adjustments. In France, for example, this weather- and seasonally-adjusted energy consumption serves as a benchmark for assessing progress toward the net-zero goal of the national strategy (*"Stratégie Nationale Bas-Carbone"*). Futhermore, all indicators used to monitor and inform future energy capacity investment decisions are expressed in weather- and seasonally-adjusted values, in line with the government's ten-year plan.[2] It is also crucial to assess the impact of a sufficiency policy such as the *"Plan de sobriété énergétique"* on energy consumption, taking into account weather and seasonal variations.[3] This is particularly relevant in the context of a change in consumer behavior due to environmental awareness and/or price increases, such as the inflationary spike triggered by Russia's invasion of Ukraine. Incorrect specification of variables and/or functional form could lead to biased estimates of energy consumption reductions and an incomplete assessment of the impact of a given policy.

[1] National emissions reported to the United Nations Framework Convention on Climate Change and the EU Greenhouse Gas Monitoring Mechanism.

[2]. The *"Programmation Pluriannuel de l'Energie"* Multi-Annual Energy Plan (MAEP) establishes the priorities for government action on energy policy for metropolitan France over the next decade, divided into two 5-year periods. The current plan covers the period from 2024 to 2033.

[3] *"Plan de sobriété énergétique"* promotes the concept of energy sufficiency, which is the voluntary and organized reduction of energy consumption. It includes various measures aimed at changing certain habits and reducing unnecessary energy consumption.
Environmental awareness has been growing in France over the years, and an annual survey by the Agence de l'Environnement et de la Maîtrise de l'Energie (ADEME) assesses a 36% awareness of climate change as the main environmental concern ADEME (2022).

The concept of Heating Degree Days (HDD) and Cooling Degree Days (CDD) has traditionally been used to explain seasonal patterns in energy consumption due to weather variations. The calculation of HDD (CDD) is based on the concept of base temperature. This indicator is only calculated if the temperature at time t is below (or above) the base temperature. The base temperature is the outdoor temperature at which agents begin to heat (cool) buildings to achieve an indoor comfort temperature. The concept of HDD (CDD) is based on the understanding that there is a non-linear relationship between temperature and energy consumption, as outlined by Henley and Peirson (1997). This methodology dates back to the 1870s when it was first used in agricultural yield studies. It is now widely used to measure adjusted energy use. While there has been extensive research on the estimation of temperature response functions for energy demand (see Fazeli et al., 2016, for a comprehensive review), a broader view of the literature (De Azevedo et al., 2015) underlines the need for a statistically validated approach, as most papers using HDD indicators still define a base temperature of 18°C, mostly based on Thom (1954).[4]

Beyond the choice of base temperature, this definition of HDD (CDD) can be challenged in three ways: first, there is no reason for the base temperature to be constant over time. For instance, Sailor and Pavlova (2003) posits that long-term climate change may markedly boost residential electricity consumption, largely due to rising air conditioning market saturation, particularly in cities with low to moderate saturation levels. This could have a far greater impact than previously estimated. Furthermore, Kennard et al. (2022) emphasises that global cooling demand is also driven by population growth in warmer regions, rather than only by a global rising of temperatures. This is evidenced by the fact that population-weighted cooling degree-days are increasing at a faster rate than area-weighted ones. This highlights the need for careful consideration of base temperature in predicting future energy demand. Consequently, these papers argue for recognising dynamic temperature patterns over extended periods, particularly in the context of climate change. Second, there is no reason for the base temperature to be constant in space, as Bessec and Fouquau (2008) underscores the imprudence of assuming a uniform temperature base across different geographical regions. Finally, there is no reason to take temperature as the sole weather indicator. Lefieux (2007) and Lundström (2017) broaden the scope to include other weather indicators in the literature, highlighting the multifaceted nature of factors influencing energy consumption, including consideration of cloud cover and the effects of wind and sunlight.

Building on this literature and relying on a statistical approach to select weather variables, we propose a new general weather indicator (GWI). We define GWI, as a vector of the optimal linear combination of heating days variable (HDV) and cooling days variable (CDV) of different weather variables among temperature, wind, sunlight duration, rain and cloudiness. To construct this indicator we proposed a two-step procedure: First, for each variable, we extract the optimal(s) threshold(s) that capture the non-linear relationship between the energy consumption and this variable using K-means. To our knowledge, this is the first time that K-means has been used to extract these thresholds. Compared to the standard likelihood-based thresholding approach, the K-means approach is more flexible: one does not need to specify a priori a functional form, between the energy consumption and the weather variable and/or the number of regimes before proceeding with the estimation. In our particular case, we also found that the K-means approach leads to

[4]See Appendix A.1 for a complete literature review.

an estimate that is more robust to spatial heterogeneity. Then, in a second step, the GWI is constructed by selecting, for each energy consumption, the optimal linear combination of these HDV and CDV using a LASSO penalization. We select the optimal specification in the sample according to different criteria: likelihood-based (AIC) and prediction error-based (RMSE). As an application of our approach, we studied electricity and natural gas consumption in France, disaggregated by delivery mode.[5] This disaggregation allows us to proxy different sectoral activities from residential and services to heavy industry. We use both monthly and daily energy consumption series: the monthly data span from 01-2012 to 12-2022 and the daily low voltage energy consumption series span from 01-2019 to 12-2021. Weather data are then provided at a fine spatial scale for daily or monthly frequency, depending on the indicator.

At the national monthly level, for the residential sector, we found that the relationship between French energy consumption and temperature is mostly L-shaped, with a heating response and a comfort zone. Contrary to the previous study, we find that the best GWI is composed of a linear combination of temperature HDV, wind HDV, and sunlight HDV. For temperature HDV, we found that the best base temperature is 15°C, independent of the introduction of other weather variables, which is neither the value currently used for French national statistics (17°C) nor the one standard in the literature (18°C). Then, regarding other economic sectors, our results found that at the monthly aggregated level, none of the CDV are significant, except for the temperature in the functional form of mean electricity consumption, highlighting the role of air conditioning in the tertiary sectors. For the industrial sector, we found that the response function to weather is in general quasi null.

We perform several checks to ensure the robustness of the GWI indicator. First, we test the robustness of the threshold estimation to spatial aggregation on residential electricity consumption: we study the 12 administrative regions of France, covering a wide range of different climatic zones. We found that our approach is robust to spatial aggregation, indeed we recover the same national threshold by averaging the one from each region, but we obtain different regional values in line with the variability of the French climatic zones. We also found the temperature CDV to be significant for residential electricity consumption in the southern region of France, highlighting the role of air conditioning in these regions. Secondly, we test the sensitivity of the GWI approach to time aggregation, still on residential electricity consumption, we study and compare daily and monthly frequency consumption. This confirms the need to include wind and sunlight duration as weather control variables on temperature for the residential sector. Then, as a by-product of our results, we estimate the parameters of the sectoral weather elasticities for French energy consumption, at the monthly and daily levels and for different regions. The estimation of these elasticities is of primary importance, as shown by Bernard et al. (2007), for example to calibrate the microsimulation models available for France (see for example Chaton, 2024; Giraudet et al., 2021; Thao Khamsing et al., 2016), leading to a more accurate response to a policy. To illustrate the importance of finding the optimal specification to adapt to weather variations, we can make a simple comparison of the estimated thermal sensitivity coefficients of residential gas consumption. It shows that

[5]Electricity and natural gas consumption account for 44% of total French final energy consumption. Both consumptions are considered in the empirical analysis of this paper, assuming that natural gas and electricity consumption react differently to weather variations. For example, it is unusual to use natural gas to cool buildings in summer. In addition, the proposed approach could be extended to other energy sources such as coal, oil, and biomass, but this is beyond the scope of this paper.

using the temperature HDV with a base temperature of 17°C instead of the optimal GWI underestimates the response of energy consumption to temperature by up to 7.3%. On average, this results in an underestimation of consumption of 0.80 TWh, equivalent to the output of over one and a half nuclear reactors per month, highlighting the critical importance of optimal weather elasticity estimates.

The remainder of this paper is organized as follows. In section (2) we present the data and associated descriptive statistics. In section (3) we present the methodological procedure. Then in section (4) we develop our main application. In section (5) we discuss further robustness analysis. Finally, a concluding section (6) is proposed.

2 Preliminary analysis of the data

2.1 Energy demand

Since one of the main goal of this paper is to provide tools for monitoring the impact of national policies to reduce energy demand, our main application focuses on the monthly aggregated gas and electricity demand time series at the French national level, initially named $[\mathbf{q}_t^{gas}]$ and $[\mathbf{q}_t^{elec}]$, respectively. The final demand of electricity is divided into three series according to the type of distribution. First, the low-voltage demand, denoted $[\mathbf{q}_t^{Low_e}]$, is defined as a response to the demand of households and small businesses, providing power to everyday appliances with a level voltage between 0.23 kV and 0.40 kV. Second, medium demand, denoted $[\mathbf{q}_t^{Med_e}]$, is described as a distribution made to facilitate the local transport of electricity to small industries, SMEs and businesses with a level tension between 15 kV and 30 kV. Thirdly, high voltage demand, denoted as $[\mathbf{q}_t^{High_e}]$, for a voltage level ranging from 63 kV to 400 kV.

Natural gas energy demand is decomposed into two series of final demand, namely: the distributed demand denoted $[\mathbf{q}_t^{Low_g}]$, defined as a network that transports gas from transmission networks to final consumers not directly connected to transmission networks. The transported demand, denoted by $[\mathbf{q}_t^{High_g}]$, denotes networks that facilitate the import of gas from terrestrial interconnections with neighbouring geographical areas and methane terminals. These monthly energy demand series can be used as a proxy for the energy demand of the three main economic sectors: Industrial, Tertiary and Residential, for easier economic interpretation, as described in Table 1.[6] This database is publicly available from 01-2000 to 12-2022 (132 observations) from the *Service des Données et Etudes Statistiques* (SDES).[7]

We also examine French residential electricity demand at daily frequency and with regional distribution.[8] These daily electricity demand series for low voltage (< 36kVA) are available from 01-01-2019 to 31-12-2023 from Enedis (1,825 observations). We checked that the aggregate daily frequency demand have the same distribution than the monthly frequency demand, at the source is not the same (see Table 17). The regional level is characterized by the study of the 12 metropolitan administrative regions of France, with the exception of Corsica: *Auvergne-Rhône-Alpes, Bourgogne-France-Comté, Bretagne,*

[6] See Table 15 for the sectoral distribution of each delivered energy type and Table 16 for the distribution of total demand across all different sectors in Appendix A.2

[7] Catalogue of energy data in France - *https://www.statistiques.developpement-durable.gouv.fr*

[8] Natural gas demand is not available at daily frequency

Table 1: Time series of energy demands by economic sector

Series		Sector	Share
Electricity - Low voltage	$q_t^{Low_e}$	Proxy for residential consumption	(80%)
Electricity - Medium voltage	$q_t^{Med_e}$	Proxy for tertiary consumption	(58%)
Electricity - High voltage	$q_t^{High_e}$	Proxy for industrial consumption	(81%)
Gas - Distributed	$q_t^{Low_g}$	Proxy for residential, tertiary and small industry consumption	(74%)
Gas - Transported	$q_t^{High_g}$	Proxy for industrial consumption, gas-fired power plant included	(94%)

Notes: Residential consumption represents 80% of the total electricity demand delivered via low voltage i.e. a 80% share.

Centre-Val de Loire, Grand Est, Hauts-de-France, Île-de-France, Normandie, Nouvelle-Aquitaine, Occitanie, Pays de la Loire, Provence-Alpes-Côte d'Azur.[9] The electricity demand of these administrative regions is made available by Enedis only from 01-01-2022 to 31-12-2023 (730 observations by region).

2.2 Weather data

Weather data are provided by the official French weather and climate service *Météo France* at a daily or monthly frequency, depending on the indicators.[10] These data are available at a fine spatial scale, as they are monitored by 539 weather stations throughout metropolitan France. In line with the literature (See for example Kennard et al., 2022), each weather station is weighted daily according to the population of the last available French census. The weighting is used to reflect the population level and the associated level of heating and cooling demand. In fact, the energy demand for a given temperature differs between countries according to their respective population densities, once again as highlighted by Kennard et al. (2022). It is important to note that the weighting by population census is particularly effective for the residential sector. However, it can result in the emergence of spatial heterogeneity on a more granular level for the tertiary and industrial sectors. For these sectors, weighting by the number of people in employment may be a more appropriate approach. However, given the sensitivity of the residential sector to weather conditions, we will maintain the population-based weighting for this general study. First, each administrative town in France is associated with the nearest weather station, i.e. a station can be associated with several towns, but each town is associated with only one station. Then, the sum of the population of each one is added to the corresponding station, giving each station a certain population weight. Finally, the daily weather indicators are weighted as follows

$$weather_s(pop) = \frac{weather_s * pop_s}{\sum_{i=1}^{s} pop_i} \tag{1}$$

Where s is one of the 539 stations in the area, pop_s is the population associated with the station and $\sum_{i=1}^{s} pop_i$ is the national population.

Based on the literature (Dell et al., 2014; Lundström, 2017), five different initial weather indicators are chosen. The cloudiness [**cloudiness**] is measured as the number of days in a month where the share of the cloudy sky is greater than or equal to 80%, the rainfall [**rain**] is measured as the amount of rainfall in 24 hours (millimeters), the temperature [**temp**] is measured as the mean temperature in 24 hours (degrees Celsius), the sunlight [**sunlight**] is measured as the sunshine duration (minutes), and finally, the

[9] Figure 11 shows the location of each administrative region on the map of mainland France.
[10] Catalogue of meteorological data in France - *https://meteo.data.gouv.fr/*

wind speed [**wind**] is measured as the daily average of 10-minute periods during which the wind speed at 10 meters above the ground is recorded (meters per second). It is important to note that the term sunshine duration is used to describe the length of time during which the ground surface is irradiated by direct solar radiation. This duration can be considered an indicator of the frequency of *favourable weather conditions*. Consequently, it not only reflects the difference in sunshine duration due to the Earth's rotation, but also takes into account days, even in summer, when solar radiation does not directly irradiate the ground surface due to overcast skies.

Table 2: Weather time-series

Series		**Units (/24h)**
Cloudiness	*cloud*	Number of days per month with overcast > 80%
Rain	*rain*	Rain level in millimeters
Temperature	*temp*	Average temperature in °C
Sunlight	*sunlight*	Duration of sunshine in minutes
Wind	*wind*	Wind speed in m/s

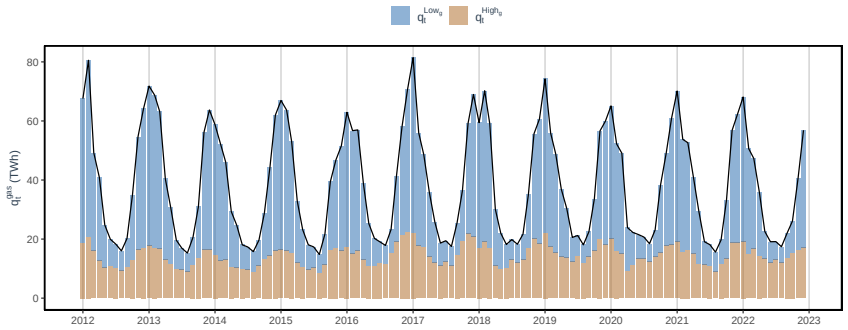
2.3 Descriptive statistics

Figure 1-(a), displays the monthly aggregated natural gas demand in France $\mathbf{q}_t^{gas}$, with the part due to the distributed $\mathbf{q}_t^{Low_g}$ and the transported $\mathbf{q}_t^{High_g}$. The data indicates a seasonal pattern in natural gas demand, primarily due to the distributed portion. Figure 1-(b), displays the monthly aggregated electricity demand in France $\mathbf{q}_t^{elec}$, with contributions from low voltage $\mathbf{q}_t^{Low_e}$, medium voltage $\mathbf{q}_t^{Med_e}$, and high voltage $\mathbf{q}_t^{High_e}$. It reveals a similar seasonal pattern, mainly driven by low voltage demand.[11] In the Appendix A.3, we show that this similar seasonal pattern appears at the daily frequency for electricity (see Figure 10 in Appendix A.3).

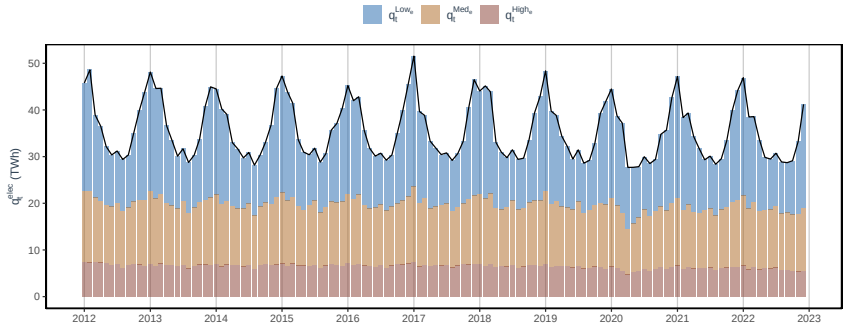
Table 3 shows the correlation between the selected energy demand time series and the five initial weather indicators. There is a strong correlation between monthly energy demand and temperature because of the large amount of energy needed to heat buildings in winter. However, other weather indicators may also correlate with energy demand. This simple correlation exercise already shows weaker correlations for cloudiness and rain. Temperature has a strong negative correlation with energy demand, about -0.96 and -0.91 for total gas and electricity demand, respectively. Sunlight has a similar relationship, with a correlation of -0.83 and -0.79 with total gas and electricity demand, respectively. However, sunlight is also strongly correlated with the level of temperature, so these indicators may carry a similar signal. Wind speed also has a significant positive correlation with energy demand at the monthly level, meaning that an increase in wind speed leads to more energy demand, which can be due to air infiltration of buildings, as highlighted in the literature (Sherman, 1987; Sinnott, 2016). The correlation with total gas and electricity demand is around 0.53 and 0.46 respectively. At the sectoral disaggregate level, the correlations show singular behavior such as: the $q_t^{High_e}$ is very weakly correlated with the climatic components, the precipitation level is only correlated with q_t^{Low} and the cloudiness is only significantly correlated with $q_t^{Med_e}$. Moreover, the correlation with temperature is still less strong for $q_t^{Med_e}$ than for the other energies and sectors. This can be explained

[11] These seasonal patterns are confirmed by the analysis of the autocorrelation function (ACF) available Figure 9 in Appendix A.3

Figure 1: Sectoral energy demand time-series



(a)



(b)

Notes : Cummulative distribution in time for electricity demand into $q_t^{Low_e}$, $q_t^{Med_e}$ and $q_t^{High_e}$ and for gas demand into $q_t^{Low_g}$ and $q_t^{High_g}$. Monthly energy demands data span 01/2012-12/2022.

by a symmetric effect between heating and cooling behavior, so the correlation is both negative and positive, the negative effect being stronger so that the overall correlation is negative (see figure 2-(b)). This result highlights the potential weakness of a correlation approach when the relationship between variable is strongly non-linear and empahsys the need for a robust methodology. Then, the weather data correlation with energy is also available at the daily frequency in Table 18 in Appendix A.3, mainly confirm the results from Table 3.

The temperature is the main driver for energy demand but, as highlighted in Table 3, the amount of sunlight or the wind speed are also correlated and could influence the demand, and should be considered in the pool of possible weather indicators that influence energy demand.

3 Methodology

In this section, we present our proposed framework for adjusting energy demand to seasonal and weather variations.

Table 3: Correlation between weather and energy time series

	temp	**sunlight**	**wind**	**rain**	**cloudiness**
$q_t^{Low_g}$	-0.969	-0.825	0.556	0.112	-0.020
	(0.000)	(0.000)	(0.000)	(0.063)	(0.739)
$q_t^{High_g}$	-0.822	-0.764	0.331	0.069	-0.057
	(0.000)	(0.000)	(0.000)	(0.252)	(0.345)
$q_t^{Low_e}$	-0.934	-0.807	0.498	0.130	0.008
	(0.000)	(0.000)	(0.000)	(0.031)	(0.896)
$q_t^{Med_e}$	-0.410	-0.324	0.031	-0.064	0.435
	(0.000)	(0.000)	(0.608)	(0.292)	(0.000)
$q_t^{High_e}$	-0.012	-0.043	0.164	0.044	-0.017
	(0.848)	(0.479)	(0.006)	(0.465)	(0.782)
q_t^{gas}	-0.966	-0.833	0.531	0.107	-0.027
	(0.000)	(0.000)	(0.000)	(0.075)	(0.656)
q_t^{elec}	-0.916	-0.790	0.466	0.100	0.134
	(0.000)	(0.000)	(0.000)	(0.097)	(0.026)

Notes : The table shows the correlation coefficients and the p-values associated. The columns are ranked according to the level of correlation.The p-values represent the probability that the null hypothesis, which represents a null correlation, is non-rejected. Thus a null p-value is interpreted as a correlation significantly different from 0. The variablesq_t^{gas} andq_t^{elec} represent the sums of the respective variables $q_t^{Low_g}$, $q_t^{High_g}$ and $q_t^{Low_e}$,$q_t^{Med_e}$,$q_t^{High_e}$. The data are at the monthly frequency and span from 01/2000 to 12/2022.

3.1 Seasonal adjustment

To construct seasonally adjusted statistics, accounting for both working-day effects and weather variations, the recommended econometric approach is based on regSARIMA methodology.[12] While the regressive component includes a set of regressors explaining both working-day effects and weather variations, the SARIMA component of the model is tailored to extract the seasonal patterns from the remaining variations in the regression stage. This model can be summarized as follows:

$$\begin{cases} q_t = \beta GWI_{q_t} + \beta_2 ly_t + \beta_3 wd_t + x_t \\ \quad \Phi(B^s)\delta(B^s)x_t = \Theta(B^s)a_t \end{cases} \tag{2}$$

Equation (2) describes the model with q_t the time series to be adjusted, ly_t is a vector of dummy variables accounting for the leap year in the specification, wd_t are two related variables accounting for working days, and GWI_t is the general weather indicator defined in section 3.2.2. This indicator may include more than one climate variable, and thus the estimated β may be a vector of parameters.[13]

The second part of the model is derived from the seasonal term of the x_t residuals and is defined by the SARIMA process, along a parameter denoted S, which takes $S = 12$ in a monthly setup. SARIMA can be decomposed into two stationary polynomials, the

[12]This approach is recommended by Eurostat. The European agency is responsible for developing, producing, and disseminating European statistics. It sets and enforces statistical standards, methods, and procedures, ensuring the production of comparable data across the European Union for various audiences. Eurostat's role is defined in Article 6 of Regulation (EC) No 223/2009 of the European Parliament and of the Council of March 11, 2009, on European Statistics.

[13]In this specification, it is worth noting that the term GWI_t is centred around the so-called Null Unified Days indicators, which is the average of past GWIs over two decades.

autoregressive $\Phi(B^s)$ and the moving average $\delta(B^s)$, and then into a nonstationary polynomial, the difference process $\Theta(B^s)$. The variations of the series that are not explained by the regressors nor by the SARIMA process are reflected in the last term a_t, in other words, a_t can be described as the adjusted time series following a $N(\mu_a, \sigma_a^2)$ distribution centered around μ_a, the mean monthly energy demand. The estimation of the model (2) is done by maximum likelihood with the function *X13* from Quartier-La-Tente et al. (2024).

3.2 Weather adjustement

Traditionally, the concepts of Heating Degree Days (HDD) (3) and Cooling Degree Days (CDD) (4) have been adopted as a solution to explain the seasonal patterns of energy demand due to weather variations. The calculation of CDD is based on the concept of base temperature, the indicator is calculated if and only if the temperature at time t is below (above) the base temperature. This method dates back to the 1870s in studies of agricultural yields and is now widely used to measure adjusted energy demand.

$$HDD = \begin{cases} T_{base} - T_t \ if \ T_t < T_{base} \\ 0 \ otherwise \end{cases} \tag{3}$$

$$CDD = \begin{cases} T_t - T_{base} \ if \ T_t > T_{base} \\ 0 \ otherwise \end{cases} \tag{4}$$

Various econometric specifications have been tested in the literature (see Fazeli et al., 2016, for a comprehensive review) to estimate the relationship between energy demand and temperature level. These approaches can be summarized as follows: the first strand of the literature focuses on linear and non-linear parametric methods. As mentioned above, the historical method is to model a linear specification with heating degree days (HDD) and cooling degree days (CDD). Initially, these models used a single balance point temperature for heating and cooling. The result is a V-shaped relationship between temperature and energy use (Mitchell, 1984) (see figure 12-(a) in the appendix A.4). For example, U.S. studies have often used 65°F (18.3°C) as the equilibrium point (Considine, 2000; De Dear and Brager, 2001; Donovan and Fischer, 1976; Pardo and Valor, 2002; Thom, 1954), while global analyses have favored 18°C (Isaac and Van Vuuren, 2009; Labriet, 2013). However, more recent studies have explored different balance point temperatures to better capture variations in energy use, using iterative methods over a range of balance point temperatures (Kissock et al., 2003; Rüth and Lin, 2006). In addition, efforts have been made to refine linear models by introducing a comfort zone (U-shape) that represents temperature ranges where no heating or cooling is required (Eskeland and Mideksa, 2010; Hekkenberg et al., 2009). (See figure 12-(b) in the appendix A.4).

Then, non-linear models have been developed to consider the complexities of consumer behaviour and heating system capacities. With these models the level of temperatures itself is used with non-linear transformation such as polynomials. For example, Henley and Peirson (1997) found a polynomial estimation to better fit data on electrical space heating compared to linear models. Asadoorian et al. (2008) and Gelegenis (2009) used log-linear formulations and polynomials to estimate temperature elasticity in electricity demand.

The second part of the literature uses a semi/non-parametric method, i.e. the relationship between energy demand and temperature is not set a priori, but is defined by the model. Early on, Engle et al. (1986) introduced a semi-parametric regression method that combines linear elements for income and energy prices with cubic and piecewise linear splines for weather variables. This approach allows flexible modeling of the temperature-energy demand relationship. Then, in a seminal work, Carcedo and Vicéns-Otero (2005) applied the Logistic Smooth Transition Regression (LSTR) model, which effectively captures the smooth response of electricity demand to temperature variations, but also provides a method for validating the temperature thresholds traditionally used. Bessec and Fouquau (2008) extended this method to analyze electricity demand in EU member states, proving that the relationship between energy demand and outside temperature is non-linear for each country in the study. Later, Hurn et al. (2016) also used LSTR modeling to asses the effects of deregulation in the electricity of market, controlling by the temperature variations.

Our econometric framework combines these two strands of the literature by introducing the new general heating weather indicator GWI_{q_t}, specific for one energy demand q_t. This indicator is defined as the vector of the optimal linear combination of the K heating days variables $\text{HDV}_{\left\{q_t, w_t^k = w_{base}^k\right\}}$. Where for $\text{HDV}_{\left\{q_t, w_t^k = w_{base}^k\right\}}$ we have $k \in 1, \ldots, K$ and $t \in 1, \ldots, T$ with T the number of observations and K one of the weather variables among temperature, wind, sunlight duration, rain and cloudiness.

Building on the standard definition of the (HDD) (3), we define a heating days variable ($\text{HDV}_{\left\{q_t, w_t = w_{base}^k\right\}}$) for one specific weather variable (w_t^k) and one energy demand q_t as:

$$HDV_{\left\{q_t, w_t^k = w_{base}^k\right\}} = \begin{cases} w_{base}^k - w_t^k \ if \ sign(\rho_{\left\{q_t, w_t^k\right\}}) * (w_t^k > w_{base}^k) \\ 0 \ otherwise \end{cases} \tag{5}$$

where w_{base}^k, extracted using K-means, is the optimal(s) extracted threshold(s) that capture the non-linear relationship between the energy demand (q_t) and one w_t^k (see Figure 2). To define one $\text{HDV}_{\left\{q_t, w_t^k = w_{base}^k\right\}}$, it is necessary to introduce the sign from the correlation between the energy demand q_t and the weather indicator w_t^k. For example for the production of HDD, the correlation between the energy demand and the temperature is negative, i.e. that lower temperature induces an increase in demand. Thus, the variable is different from zero if $-(w_t^k > w_{base}^k)$, in other terms, if $(w_t^k < w_{base}^k)$. At the opposite, for the production of the wind HDV, the correlation between the energy demand and the wind is positive i.e. that stronger wind speed induces an increase in demand. Thus, the variable is different from zero if $(w_t^k > w_{base}^k)$. The literature, highlights the role of the air-conditioning, leading for a V shape between the demand q_t and the temperature (see Mitchell, 1984). To test the good functional form (V, U or L shape see Figures 12 in Appendix A.4), we enhance our GWI, to take into account this cooling effect. With similar notation of (5), we defined the CDV as follows:

$$CDV_{\left\{q_t, w_t^k = w_{base}^k\right\}} = \begin{cases} w_t^k - w_{base}^k \ if \ sign(\rho_{\left\{q_t, w_t^k\right\}}) * (w_t^k > w_{base}^k) \\ 0 \ otherwise \end{cases} \tag{6}$$

It is important to stress that, w_{base}^k in (5) can be different from (6), depending on the functional form.

The next subsections present the K-means procedure to define potential thresholds for one couple of q_t and w_t^k and then the LASSO penalisation that selects the optimal linear combination of both HDV and CDV for one q_t.

3.2.1 Numbers of regimes and threshold extraction

One of the most used approaches in the literature to define the energy demand response function to temperature is drawn upon Carcedo and Vicéns-Otero (2005) and Bessec and Fouquau (2008), which employs a Smooth Transition AR (STAR) method based on a predetermined definition of the data generation function. However, a smooth method induces an a priori definition of a function that best describes the relation between the series. One can also use, a non-smooth methodology grounded in the method of multivariate Threshold Vector Autoregression (TVAR) from Lo and Zivot (2001), based on the Self-extracting AR (SETAR) modelling (Chan et al., 1985). However, estimating thresholds in this setup is prone to sensitivity towards outliers or noise present in the data, which can lead to potentially biased or inconsistent estimates. Incorrect specification of the number of thresholds may result in misspecified models that fail to accurately capture the genuine underlying dynamics of the data. Empirically, it is uncommon to define more than three regimes with the associated two thresholds in a TVAR modelling because it becomes computationally heavy: every time a regime is added it strongly increases the size of the grid search. Moreover, in this context of the relationship between climate and energy demand, imposing a predetermined number of regimes a priori may overlook the true complexity of the relationship and means relying directly on visual representation and the existing shape of the relationship (such as V-shaped or U-shaped).[14] This underscores the need for a thorough and flexible approach to the model specification that allows for the discovery of the most appropriate response functional form and the number of regimes to accurately characterize this relationship. To do so, our approach relies on unsupervised classification and in particular using time-series clustering methods.

More formally, our time-series clustering can be defined as follows: given our dataset of a vector of two time-series containing q_t, and one of the n weather variables denoted by w_t^k. We define $D = \left\{q_t, w_t^k\right\}$, and we find an unsupervised partitioning of D into $C = \{C_1, C_2, \ldots, C_N\}$, where N is a hyperparameter correspond to the number of clusters that you need to set a priori. We make this partitioning in such a way that homogeneous time series are grouped based on a certain similarity measure (See Liao, 2005; Aghabozorgi et al., 2015, for a comprehensive review on the definition and application of clustering with time series.) [15]

Drawing on the seminal K-means algorithm (MacQueen, 1967), we operate iteratively assigning each data point to the nearest cluster centroid based on the Euclidean distance metric. Subsequently, the centroids are recalculated as the mean of the data points assigned to each cluster. This iterative process continues until convergence, resulting in clusters that exhibit similar characteristics or behaviours (see Algorithm 1).

[14] In appendix B.2, we perform a robustness analysis using the TVAR methodology. We find our approach is robust at the French national level, extracting similar thresholds. However, moving to the desegregated demand by regions, the heterogeneity of climate leads to the apparition of a cooling behaviour and thus, for a constant number of regimes, the extracted threshold using the TVAR methodology is biased by the cooling behaviour.

[15] Our application of clustering on time-series is similar to the Market Regime Clustering Problem (MRPC) in Finance (Horvath et al., 2021), MRPC involves segmenting returns into different groups or regimes, each characterized by distinct underlying distributions. Another close methodology is used by Greevy et al. (2024) to detect regime switching in finance time-series, but the clustering algorithm is used on statistics distribution variables rather than the series themselves.

Algorithm 1: k-means algorithm

Data: Set of T data points $D = \{q_t, w_t^k\}$ in $\mathbb{R}^2$, number of clusters N
Result: N clusters $C = \{C_1, C_2, \ldots, C_N\}$

Initialization: Choose N initial cluster centroids randomly: $C = \{c_1, c_2, \ldots, c_N\}$;
repeat
  for each data point $d_t = (q_t, w_t^k)$ **do**
    Assign d_t to the nearest cluster centroid:

$$\arg\min_{c_j \in C} \|d_t - c_j\|^2$$

  end
  Recalculate the cluster centroids as the mean of the data points assigned to each cluster:;
  for each cluster C_j **do**
    $c_j = \frac{1}{|S_j|} \sum_{d_t \in S_j} d_t$ where S_j is the set of data points assigned to cluster C_j;
  end
until convergence;

One challenge in using the K-means algorithm is the requirement to pre-assign the hyperparameters, the N number of clusters, which may not always be readily available or feasible to determine in real-world applications. This limitation can lower the algorithm's ability to produce natural clustering results and is recognized as one of its drawbacks. However, a potential solution, particularly when the data dimensionality is low, is the Within Sum of Square (WSS) criterion (7). The WSS measures the compactness of clusters by summing the squared distances between each data point and its assigned cluster centroid. This criterion typically decreases as the number of clusters increases, indicating tighter cluster formations. However, beyond a certain point, further increasing the number of clusters may yield diminishing returns in terms of reducing the WSS. To address this, a common strategy involves identifying the number of clusters N where the rate of decrease in the WSS significantly slows down. This inflexion point, often referred to as the "elbow" point, serves as a practical guide for selecting an appropriate value for N, striking a balance between maximizing cluster compactness and minimizing model complexity.[16]

$$WSS_N = \sum_{j=1}^{N} \sum_{t=1}^{n_N} ||d_t - \bar{cj}||^2 \tag{7}$$

$$sWSS = \frac{WSS_{N+1}}{WSS_N} - 1 \tag{8}$$

To let the algorithm discover the number of regimes to accurately characterize the relationship, we first compute the WSS criterion for a set of 10 different numbers of N clusters ($N \epsilon [1; 10]$). Once the WWS is computed for each number of clusters, we compute the selection Within Sum of Square (sWSS) (8) that asses the variation of compactness between the number of clusters. We set a subjective threshold, such that the number N of

[16] It should be noted that no cross-validation process is used when dealing with our time series data as it could lead to select unbalanced mean temperature patterns such as only months with high temperature which correspond to the summer season or inversely only months with low temperature which correspond to the winter season.

the cluster selected is the last N number that leads to an increase in compactness greater than 60% compared to the previous number of clusters.[17] Once the number of clusters, which can be interpreted as the number of regimes, is defined then the thresholds can be extracted. The extracted thresholds (w_{base}) are defined as the mean maximum value of (w_t) for each cluster. The means from the top $\alpha\%$[18] of the weather variable (w_t^k) is computed rather than the direct maximum to avoid the potential bias from a one-time outlier. Once we obtain (w_{base}^k), we construct $HDV_{\left\{q_t, w_t^k = w_{base}^k\right\}}$ using (5), and we repeat this step for every w_t^k, in all the weather variable. Similarly, we use the same procedure to find w_{base}^k for $CDV_{\left\{q_t, w_t^k = w_{base}^k\right\}}$ using (6).

3.2.2 Penalization procedure to construct the GWI

Once the pool of all the K weather variable-related HVD and CDV indicators is constructed, a variable selection process is employed through penalization. Penalization models incorporate a hyperparameter denoted as λ within the error term minimization process, to select the optimal linear combination among the regressors.

$$\begin{cases} q_t = \beta' X_t + \varepsilon_t \\ \min \sum^t (q_t - \hat{q}_t)^2 + \lambda ||\beta||_k \end{cases} \tag{9}$$

Where q_t, is one energy demand variable of T observations, $\beta = [\beta_1, \cdots, \beta_{2*K}]$ is $2 * K \times 1$ vector of the estimated coefficient of our linear penalization regression, and $X_t = \left[HVD_{q_t, w_t^1 = w_{base}^1}, \cdots, HVD_{q_t, w_t^K = w_{base}^K}, CVD_{q_t, w_t^1 = w_{base}^1}, \cdots, CVD_{q_t, w_t^K = w_{base}^K}\right]$ is a $2 * K \times T$ regressor weather vector, containing the HDV and the CDV.[19] ε_t is the residuals vector following a $N(0, \sigma^2)$, where σ^2 is variance matrix of the residuals.

The penalisation model can be described as in equation (9) with λ the level of penalization, the higher $\hat{\lambda}$ the more penalized the model, the lower the $\hat{\lambda}$ the less penalized the model. $||\beta||_k$ reflects the norm, which is defined as the LASSO (l_1) norm in our approach, q_t denotes the initial seasonal time series and HDV_t is the weather indicator. For robustness purposes, at each iterative step, we run 100 batches of estimation by varying the lambda value in a range between 0.001 and 1 to test different levels of penalty. Then the mean of the 100 estimations is computed for each variable of the initial pool.

Denoted the optimal sparse penalized β^*, containing some zeros for the weather variables that are not selected by the LASSO procedure, we defined the GWI_{q_t} as:

$$GWI_{q_t} = \left[1_{\beta_1^* \neq 0} HVD_{q_t, w_t^1 = w_{base}^1}, \cdots, 1_{\beta_K^* \neq 0} HVD_{q_t, w_t^K = w_{base}^K}, 1_{\beta_1^* \neq 0} CVD_{q_t, w_t^1 = w_{base}^1}, \cdots, 1_{\beta_K^* \neq 0} CVD_{q_t, w_t^K = w_{base}^K}\right] \tag{10}$$

where $1_{\beta_k^* \neq 0}$, is a dummies variable taking the value 1, if the corresponding HVD_{q_t, w_t^k} is selected and 0 if is not.

[17] The choice of 60% is made so that the increase of a cluster leads to a significant decrease in the inertia. Figure 13 in appendix A.5 showcases the inertia variation in terms of number of clusters for one of the energy demand series studied.

[18] The α can be different regarding the frequency of the data studied.

[19] One could also include all the interactions term between different HDV (or CDV). In our empirical illustration, none of these terms is selected, we decide for the sake of clarity not to mention that in the methodology part.

4 Empirical results

One of the goals of this paper is to develop tools for monitoring the impact of national policies aimed at reducing energy demand. Our primary application focuses on the analysis of a monthly database from January 2012 to December 2022. This extensive dataset allows us to study the general behavior of energy demand in relation to weather variations over a ten-year period.

4.1 A new baseline temperature

To illustrate how our proposed K-means-based approach estimates the optimal w_{base} for an HDV_{q_t} variable with respect to a demand q_t, we first examine the relationship between the sectoral energy demands and the temperature variable. Figure 2 shows the relationship between ordered sectoral energy demand and ordered temperature, for natural gas and electricity. In this figure, the different dashed lines represent the extracted base thresholds w_{base} reported in Table 4, and the different colours represent the regimes defined by the clustering algorithm.

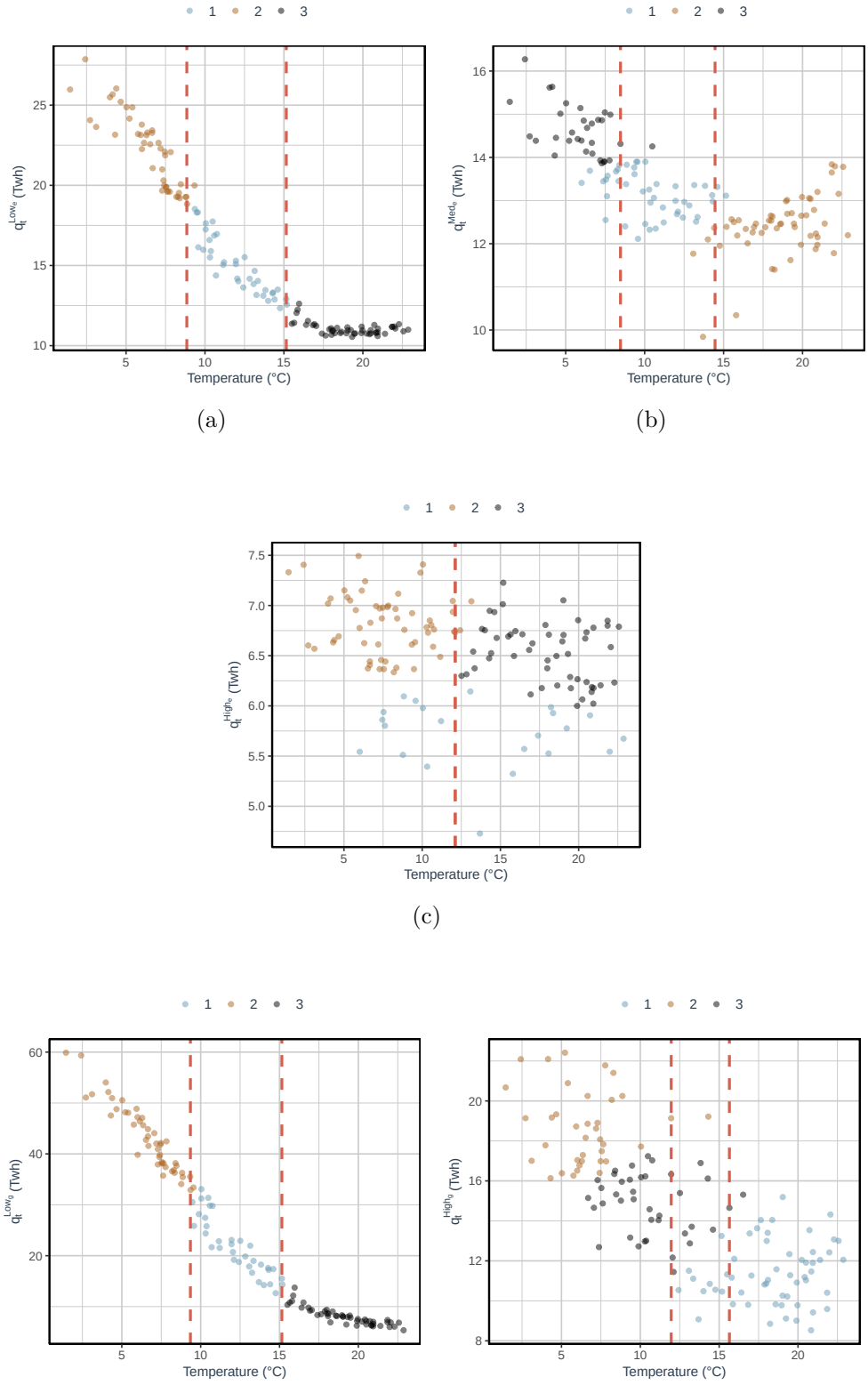
Table 4-(a), shows the value for the HDV related to temperature threshold is on average 15°C, for all the energy demand which is different from the thresholds that are commonly used in France for official statistics adjustment (17°C), the most commonly used in the literature (18°), or the recommended one by EUROSTAT.[20] We will consider, these three HDVs denoted by $HDV_{temp=17}$, $HDV_{temp=18}$, and $HDV_{temp=EU}$, as our benchmarks for the further analysis developed in the rest of this section.

As explained in section 3.2.2, we then select a variable based on the LASSO penalization methods. Table 5 showcases the output from this process, i.e. the optimal linear combination of HDV_{temp} and CDV_{temp}, that is selected in the composition of the GWI_{q_t} in equation (10), for each energy demand q_t. Table 5-(a), confirms that 15°C is the optimal threshold and that none of the three benchmarks is selected.

[20] We define $HDV_{temp=EU}$ as the one recommended by EUROSTAT, constructed as:

$$HDV_{temp=EU} = \begin{cases} T_{18} - T_t \; if \; T_t < T_{15} \\ 0 \; otherwise \end{cases} \tag{11}$$

Figure 2: Clustering analysis of monthly energy response to temperature



Notes : (a) $q_t^{Low_e}$ (b) $q_t^{Med_e}$ (c) $q_t^{High_e}$ (d) $q_t^{Low_g}$ (e) $q_t^{High_g}$. These figures are a visual representation of the K-means algorithm results. The dashed lines represent the regime-switching point, or base temperature and the clusters represent the different regimes. Monthly energy demands and weather data span 01/2012-12/2022.

Table 4: Estimated HDV_{base} for monthly energy demands based on clustering

	Temperature (a)					
	$q_t^{Low_g}$	$q_t^{High_g}$	$q_t^{Low_e}$	$q_t^{Med_e}$	$q_t^{High_e}$	$\hat{w_{base}}$
$w_{base:temp1}$	9.34	11.95	8.84	8.47	-	9
$w_{base:temp2}$	15.15	15.65	15.15	14.46	12.11	15
	Sunlight, Wind, Rain and Cloudiness (b)					
	$q_t^{Low_g}$	$q_t^{High_g}$	$q_t^{Low_e}$	$q_t^{Med_e}$	$q_t^{High_e}$	$\hat{w_{base}}$
$w_{base:wind1}$	3.09	3.08	3.09	-	-	3
$w_{base:wind2}$	3.48	3.28	3.59	3.40	3.31	3.5
$w_{base:sunlight1}$	257	294	282	290	314	300
$w_{base:sunlight2}$	412	428	412	401	569	400
$w_{base:rain1}$	1.72	-	1.94	-	-	2
$w_{base:rain2}$	2.37	3.43	1.98	2.78	2.37	2
$w_{base:cloud1}$	23.54	24.59	23.15	23.47	23.40	24

Notes : Extracted thresholds w_{base} using the clustering procedure. Monthly energy consumptions and weather data span 01/2012-12/2022.

Table 5: LASSO estimates for monthly energy demands

	Temperature (a)				
	$q_t^{Low_g}$	$q_t^{High_g}$	$q_t^{Low_e}$	$q_t^{Med_e}$	$q_t^{High_e}$
$CDV_{temp=15}$	0,00	0,00	0,00	0,03	0,00
$HDV_{temp=9}$	0,00	0,00	0,00	0,00	0,00
$HDV_{temp=15}$	**0,88**	**0,37**	**0,94**	**0,72**	**0,40**
$HDV_{temp=17}$	0,00	0,00	0,00	0,00	0,00
$HDV_{temp=18}$	0,53	0,24	0,70	0,00	0,00
$HDV_{temp=EU}$	0,00	0,00	0,00	0,00	0,06
	Sunlight, Wind, Rain and Cloudiness (b)				
	$q_t^{Low_g}$	$q_t^{High_g}$	$q_t^{Low_e}$	$q_t^{Med_e}$	$q_t^{High_e}$
$HDV_{wind=3}$	**1,25**	**0,45**	**1,37**	**0,47**	0,00
$HDV_{wind=3.5}$	0,00	0,00	0,00	0,00	**0,26**
$HDV_{sunlight=300}$	0,00	0,00	0,00	0,00	0,00
$HDV_{sunlight=400}$	**0,12**	**0,37**	**0,06**	**0,16**	**0,17**
$HDV_{rain=2}$	**0,07**	0,00	**0,08**	0,00	0,00
$HDV_{cloud=24}$	0,00	0,00	0,00	**0,03**	**0,13**

Notes: If the estimates is less than |0.01| the variable is consider as not significant and put to 0. Monthly energy consumptions and weather data span 01/2012-12/2022.

4.2 The role of the other weather variables

To highlight the role of the other weather variable, we estimated different w_{base} for all listed weather variables (w_t) in our dataset (see section 2 for the full list) using the GWI procedure from section 3. Table 4-(b) shows the estimated threshold for all considered variables, i.e. cloudiness, rain, sunlight and wind speed.[21] Then, our penalty selection procedure shows potential HDV for mainly the wind speed and the sunlight duration and then the rain or cloudiness for certain energy demands, as shown in Table 5 -(b). For the link between cloudiness and electricity demand in the industrial sector, it should be noted that the linked is reversed as described by Koenig (1979), i.e. that an increase in the demand for electricity leads to an increase in industrial production and thus in the share of the sky overcast by clouds, this relationship is describe by figure 18-(c).

After constructing the GWI_{q_t} for each q_t, we analyze the output of the seasonal regression process, more precisely, the main output policymakers are interested in is the β in (2) in section 3.1, indeed these correspond to the weather elasticities for energy demand. In Table 6 we summarize the obtained estimated $\hat{\beta}$-weather coefficients associated with each GWI_{q_t}. These coefficients are expressed in TWh/HDV, for example for the HDV_{temp} based on the temperature it can be interpreted as the increase of energy demand in TWh when the HDV_{temp} increases of 1°C, *ceteris paribus*.[22]

The specification evaluation is made through two criteria: the Akaike Information Criterion (AIC) and the sum of squared residuals. In the set-up of seasonal adjustment, it is less feasible to use metrics that rely on in and out samples since a priori the target estimation of the adjusted time series is not known. However Cui et al. (2023) highlighted that AIC model selection is consistent with weather study if one of the specifications under consideration nests the true one. Since our methodology induces two preliminary steps to define an initial pool of robust indicators, the hypothesis is made that one of our specifications nested the true data-generating process. The AIC, functioning as a penalizing criterion, guides the selection process, privileging the in-sample specification with the lowest AIC. The Root Mean Square Error (RMSE) is used to oversee the phenomenon of overfitting inherent in the modelling exercise.

The first striking result in Table 6 is that the introduction of these HDV, constructed with w_t different from temperature, induces a better specification for any energy demand that shows a significant response to weather variations. In other words, except for $q_t^{High_e}$, all the estimates are considered less biased when we introduce our GWI_{q_t}, since it allows to reduce both the AIC and the RMSE metrics. In addition to being significant, the introduction of these variables leads to a change in the estimation of the thermal sensitivity coefficient. For example, for $q_t^{Low_g}$ the thermal sensitivity is initially estimated to be 0.1037 TWh/$HDV_{temp=15}$ and after the introduction of wind speed and sunlight duration it is estimated to be 0.1044 TWh/$HDV_{temp=15}$. This is a difference of 0.7 GWh/$HDV_{temp=15}$. In other words, the sensitivity to temperature for $q_t^{Low_g}$ is underestimated when only temperature is considered.

However, the response function differs according to the energy and the sector of demand. Focusing on $q_t^{Low_g}$, Table 6-(5) showcases a significant response to $HDV_{wind=3.5}$

[21] Figures showing the relationship between the ordered energy demand and the considered weather variable are available in Appendix B.3

[22] For every energy, a first specification without any weather regressors is estimated, to prove the need to adjust the seasonality from the weather variations.

and $HDV_{sunlight=400}$, that can respectively be interpreted as the variation of $q_t^{Low_g}$ in TWh when the wind speed increase of one unit and when the sunlight duration decrease of one unit. Both estimators are positive, indicating that an increase in wind speed or a decrease in sunlight duration leads to an increase in $q_t^{Low_g}$ demand. However, only the response to wind speed is significant at the national aggregate level, and not the sunlight duration that which however seems to have a role to play for residential demand. This result is likely due to the frequency and territorial aggregation of data for this main study. Indeed, section 5.2 presents a similar estimation but with data at a daily frequency and the sunlight duration is estimated as statistically significant for $q_t^{Low_e}$. Table 6-(10) presents a significant estimate for the response function to HDV wind speeds of 3. For both $q_t^{Low_e}$ and $q_t^{Med_e}$, the model estimates a significant response function to $HDV_{wind=3}$, although to a lesser extent than for q_t^{gas} demands.

These results lead to a particular interpretation for the residential sector, where for both gas and electricity, it is the perceived temperature rather than the measured temperature that explains variations in demand. However, the response to this perceived temperature is different, with a higher elasticity for gas demand than for electricity demand, in line with the existing literature (see for exemple Alberini et al., 2011)

Table 6: Monthly estimates of energy demand response to weather sensibility

$q_t^{Low_g}$	(1)	(2)	(3)	(4)	(5)	
$HDV_{temp=17}$	(-)	0.0973***	(-)	(-)	(-)	
		(45.93)				
$HDV_{temp=EU}$	(-)	(-)	0.0907***	(-)	(-)	
			(39.62)			
$HDV_{temp=15}$	(-)	(-)	(-)	0.1037***	0.1044***	
				(48.07)	(49.78)	
$HDV_{wind=3}$	(-)	(-)	(-)	(-)	0.0270*	
					(2.16)	
$HDV_{sunlight=400}$	(-)	(-)	(-)	(-)	0.0002	
					(1.12)	
$HDV_{rain=2}$	(-)	(-)	(-)	(-)	0.0046	
					(0.99)	
AIC	698	348	378	334	321	
RMSE	3.91	0.92	1.04	0.85	0.78	
$q_t^{High_g}$	(6)	(7)	(8)	(9)	(10)	
$HDV_{temp=17}$	(-)	0.0223***	(-)	(-)	(-)	
		(10.50)				
$HDV_{temp=EU}$	(-)	(-)	0.0209***	(-)	(-)	
			(10.13)			
$HDV_{temp=15}$	(-)	(-)	(-)	0.0241***	0.0235***	
				(11.20)	(11.34)	
$HDV_{wind=3}$	(-)	(-)	(-)	(-)	-0.0293**	
					(-2.66)	
$HDV_{sunlight=400}$	(-)	(-)	(-)	(-)	-0.0001	
					(-0.45)	
AIC	449	372	376	366	359	
RMSE	1.40	1.00	1.02	0.97	0.92	
$q_t^{Low_e}$	(11)	(12)	(13)	(14)	(15)	
$HDV_{temp=17}$	(-)	0.0317***	(-)	(-)	(-)	
		(36.49)				
$HDV_{temp=EU}$	(-)	(-)	0.0300***	(-)	(-)	
			(33.07)			
$HDV_{temp=15}$	(-)	(-)	(-)	0.0341***	0.0342***	
				(40.28)	(39.07)	
$HDV_{wind=3}$	(-)	(-)	(-)	(-)	0.0085	
					(1.75)	
$HDV_{sunlight=400}$	(-)	(-)	(-)	(-)	0.0001	
					(0.22)	
$HDV_{rain=2}$	(-)	(-)	(-)	(-)	0.0036	
					(1.90)	
AIC	430	133	155	111	104	
RMSE	1.27	0.35	0.38	0.32	0.31	
$q_t^{Med_e}$	(16)	(17)	(18)	(19)	(20)	(21)
$HDV_{temp=17}$	(-)	0.0072***	(-)	(-)	(-)	(-)
		(12.62)				
$HDV_{temp=EU}$	(-)	(-)	0.0069***	(-)	(-)	(-)
			(12.54)			
$HDV_{temp=15}$	(-)	(-)	(-)	0.0079***	0.0083***	0.0086***
				(13.71)	(14.73)	(18.47)
$CDV_{temp=15}$	(-)	(-)	(-)	(-)	0.0050***	0.0062***
					(5.61)	(7.27)
$HDV_{wind=3}$	(-)	(-)	(-)	(-)	(-)	0.0106***
						(4.04)
$HDV_{sunlight=400}$	(-)	(-)	(-)	(-)	(-)	0.0001
						(1.42)
$HDV_{cloud=24}$	(-)	(-)	(-)	(-)	(-)	0.0094
						(0.27)
AIC	141	41	41	33	2	-19
RMSE	0.38	0.25	0.25	0.24	0.21	0.19
$q_t^{High_e}$	(22)	(23)	(24)	(25)	(26)	
$HDV_{temp=17}$	(-)	0.0020***	(-)	(-)	(-)	
		(7.00)				
$HDV_{temp=EU}$	(-)	(-)	0.0019***	(-)	(-)	
			(7.09)			
$HDV_{temp=15}$	(-)	(-)	(-)	0.0021***	0.0021***	
				(7.19)	(6.97)	
$HDV_{wind=3.5}$	(-)	(-)	(-)	(-)	-0.0001	
					(-0.05)	
$HDV_{sunlight=400}$	(-)	(-)	(-)	(-)	0.0000	
					(0.91)	
$HDV_{cloud=24}$	(-)	(-)	(-)	(-)	-0.0030	
					(-0.81)	
AIC	-61	-102	-103	-103	-99	
RMSE	0.168	0.142	0.141	0.141	0.139	

Notes: The estimates are from a regSARIMA regression with monthly energy consumption level in TWh as the dependent variable. The table presents the estimates for weather components that significantly impact energy consumption. However. the model also estimates fixed SARIMA components. i.e. the cyclic elements. and fixed control variables for business days: leap year. weekend and holidays effects. For the SARIMA component. large order of polynomials δ. Φ and Θ are fix along the defined using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) (see Figure 9). thus the process is defined as (1.0.1)(0.1.1). Monthly energy consumptions and weather data span 01/2012-12/2022.

4.3 Economic implications

This section presents two principal economic implications of the empirical results. Firstly, the shape of the relationship between energy demand and weather variables is analysed. This contributes to a body of literature that characterises consumer behaviour in response to temperature variations across a year (see for exemple Engle et al., 1986; Dubin, 2008). This provides insights into the need for heating and/or cooling, as well as the existence of a so-called comfort zone where energy use is not needed to regulate building temperature. We will then discuss the impact on the estimated weather elasticity, which is crucial for policymakers and energy providers as it informs strategies for energy supply management and infrastructure planning (see for example Chaton, 2024; Giraudet et al., 2021; Thao Khamsing et al., 2016). By quantifying the manner in which energy demand responds to temperature fluctuations, it facilitates the development of more accurate predictive models, which can be used to mitigate the effects of extreme weather events on energy systems (see Sgarlato and Ziel, 2023). In conclusion, our findings have significant implications for energy policy, climate adaptation strategies and economic planning.

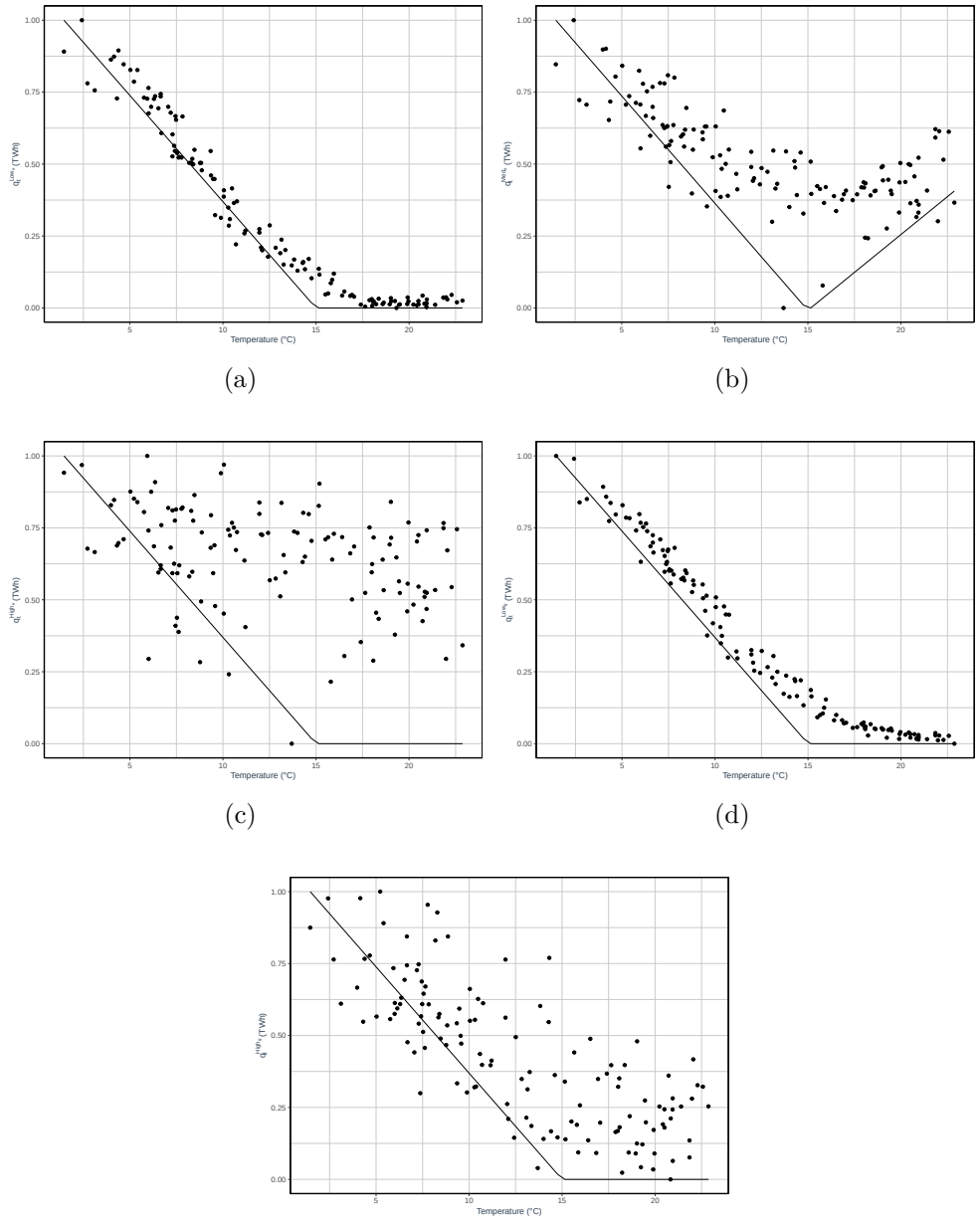
Fonctional form of France sectoral energy demand

One of the main interests is the response of energy demand to the temperature, we mainly focus on the HDV_{temp}.[23] First looking at figure 3 - (c), the relationship between $q_t^{High_e}$ and w_t^{temp} is almost constant which confirms the non-correlation between the two variables, already exhibited in Table (3). Thus the estimated thermo-sensitivity on Table 6 is the smallest one: 0.0021 TWh/$HDV_{temp=15}$. Second, Figures 3-(a), 3-(d) and 3-(e) showcase a response function to the temperature that can be defined as an L-shaped with one balance base temperature and a comfort regime with no response to a variation in temperature. Indeed, from Figures 2-(a), 2-(b) and 2-(d) with the clustering algorithm results, it presents an initial response function to temperature with two heating regimes and a comfort zone and thus two balance points: 9°C and 15°C (see Table 4), in line with the functional form for a residential panel defined by Dubin (2008). In these figures (2-(a), 2-(b) and 2-(d)), the sensitivity appears to be lower between 9°C and below, indicating that a decrease of one degree Celsius results in a lowest increase in energy demand. The primary assumption regarding this behaviour is that agents reach their budget constraint after a certain amount has been spent on heating and are thus less inclined to demand more energy. However, the selection process output in Table 5 put forward that the balanced point at 9°C is not as statistically significant as a balance point at 15°C. Table 6 shows that the residential sector is the one with the strongest thermo-sensitivity respectively 0.1044 TWh/$HDV_{temp=15}$ and 0.0342 TWh/$HDV_{temp=15}$ for $q_t^{Low_g}$ and $q_t^{Low_e}$. Third, looking at Figure 3-(b), the temperature response function from $q_t^{Med_e}$ can be described as a classical V-shaped with a single balance base temperature at 15°C. This shape has been described by Mitchell (1984) and corresponds to a single balance point between heating and cooling behaviour. From Figure 2 - (b) with the clustering algorithm results, it presents an initial response function to temperature with two heating regimes and a cooling regime and two associated balance points: 9°C and 15°C (see Table 4). However, the selection process output in Table 5 once more put forward that the balanced point at 9°C is not as statistically significant as the balance

[23] Figures showing the function response to the other significant weather variable are available in Appendix B.4

point at 15°C. This Table also allows to assess the first significance of the introduction of a CDV based on temperature for the response function to temperature from $q_t^{Med_e}$. This thermo-sensitivity to high temperature is confirmed by the estimation in Table 6 where the estimated coefficient is 0.0061 TWh/$CDV_{temp=15}$ while the estimation for the sensitivity to low temperature is of 0.0086 TWh/$HDV_{temp=15}$.

Figure 3: Linear normalized optimal estimation responses to temperature



Notes : (a) $q_t^{Low_e}$ (b) $q_t^{Med_e}$ (c) $q_t^{High_e}$ (d) $q_t^{Low_g}$ (e) $q_t^{High_g}$ - These figures are a normalised visual representation of the linear estimations of thermo sensitivity for the different energy demand. This representation only allows to take into consideration the shape of the relationship but does not correspond to the exact fitted/estimated data on the regression process. Monthly energy demands and weather data span 01/2012-12/2022.

Weather elasticities estimate

To illustrate the importance of finding the optimal specification to adapt to weather variations, we can make a simple comparison of the estimated thermo-sensitivity coefficients between Table 6-(2) and Table 6-(5), as reported in Table 7. For $q_t^{Low_g}$, the difference results in a 7.3% decrease in thermosensitivity, which means that using $HDV_{temp=17}$ instead of the optimal $GWI_{q_t^{Low_g}}$ underestimates up to 7.3% of thermosensitivity, and this fraction of energy variability could be incorrectly explained by other factors such as price and socioeconomic components, or remain as unexplained residuals. For example, December 2015 was the warmest December in the 2012-2022 period, resulting in negative HDV_{temp} relative to the average December HDV_{temp}. Thus, for December 2015, using the elasticity based on $HDV_{temp=17}$ gives an estimate of energy demand due to temperature of -11.06 TWh. This demand is underestimated by 7.3% compared to the use of GWI_{q_t} which estimates -11.86 TWh. There is a decrease in demand of 0.80 TWh that is not due to weather variations if the weather vector (GWI_{q_t}) is not properly optimized. 0.8 TWh also corresponds to the production of more than one and a half nuclear reactors per month in France (0.5TWh), underlining the role of optimal weather elasticity estimates for infrastructure planning.

Table 7: Decrease in the estimated thermo-elasticities between GWI_{q_t} and $HDV_{temp=17}$

Sectors	**Decrease**
$q_t^{Low_g}$	-7.3 %
$q_t^{High_g}$	-5.4 %
$q_t^{Low_e}$	-7.2 %
$q_t^{Med_e}$	-19.4 %
$q_t^{High_e}$	-5.0 %

5 Robustness analysis

This section provides a robust analysis of our GWI approach to, time subsample, daily and spatial data, considering 12 administrative regions of France.[24]

5.1 Rolling time window analysis

Section 4.1 only presents the penalisation process for the monthly aggregated level from 01/2012-12/2022, thus a rolling exercise of ten years moving windows from 2000 to 2022 allows an overview of the variations of the optimal composition of the GWI_{q_t} index for each q_t considered.

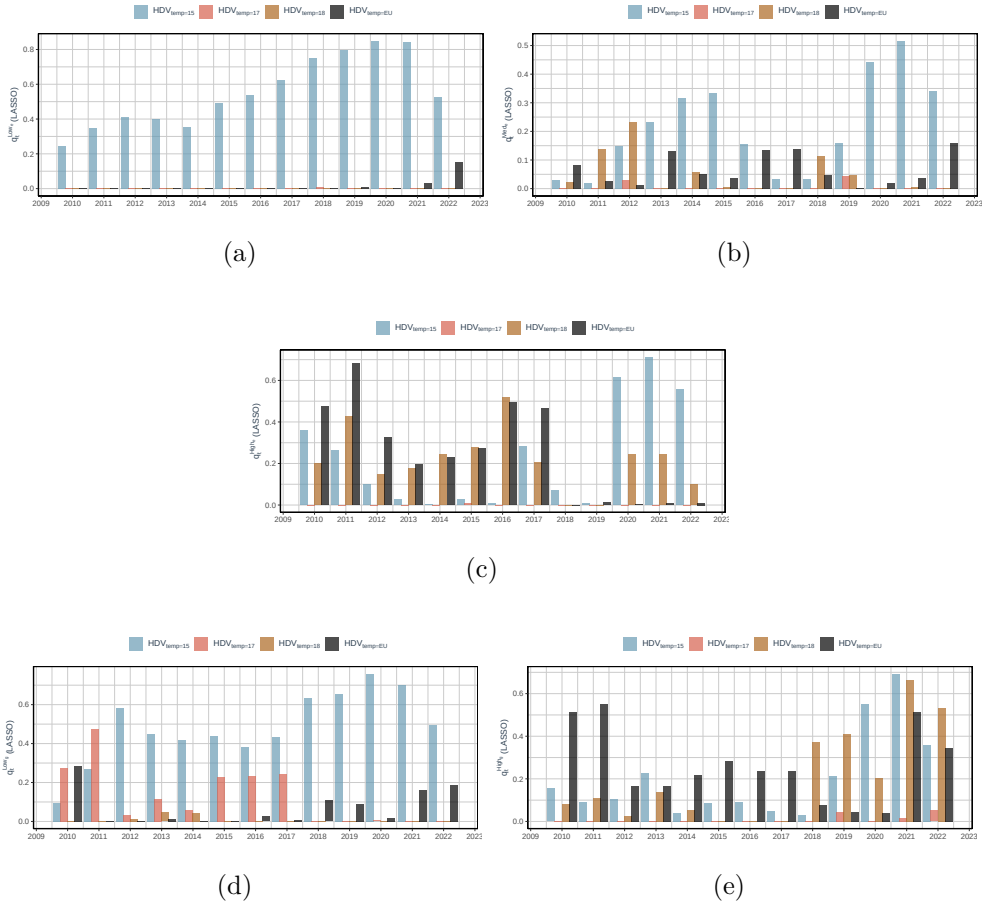
Figure (4) illustrate the outcomes of the rolling LASSO exercise concerning the HDV_{temp} selection. For a base temperature of 17°C, the rolling analysis suggests this temperature as potentially optimal for $q_t^{Low_g}$ during the 2000-2011 period. The LASSO algorithm does not penalize this variable to zero until 2018, indicating its sustained relevance over time.

[24]The 12 administrative regions considered are the metropolitan administrative regions, without Corsica that is not served by ENEDIS for $q_t^{Low_e}$.

In contrast, for a base temperature of 18°C, the rolling exercise identifies it as a significant base temperature for both $q_t^{High_g}$ and $q_t^{High_e}$. Nevertheless, the demand pattern linked to this temperature exhibits a weaker correlation with temperature variations, as demonstrated in Table 3. This reduced robustness in identifying an optimal base temperature elucidates why 18°C is selected for $q_t^{High_g}$ in Table 5.

This section emphasizes the need to tailor the threshold to the specific sub-sample being studied, as different time windows could result in different behavioral responses to temperature. This could raise concerns about the comparability of the policy evaluation across different time-varying subsamples. To illustrate, a substantial policy investment in energy efficient buildings should result in a shift to a lower baseline temperature due to increased building insulation.

Figure 4: HDV_{temp} LASSO estimates on time-varying window



Notes : (a) $q_t^{Low_e}$, (b) $q_t^{Med_e}$ and (c) $q_t^{High_e}$ (d) $q_t^{Low_g}$ and (e) $q_t^{High_g}$ - Data span the period 2000-2022, one month added at each rolling step.

5.2 Daily data

To study the robustness of our approach to higher frequency, we applied the same procedure as in section 4 to the daily data, to decompose the optimal base temperature and possible other weather regressors. Recall that at the daily frequency, *only residential electricity demand* is available.

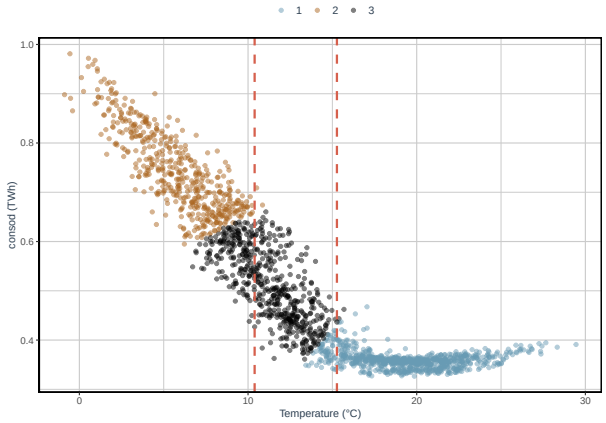
We present the first step using our K-means procedure and it leads to defining once again a base temperature of 15°C (see Table 8) and Figure 5 that showcases the non-linear link between energy demand and temperature is similar to the estimated one using the monthly dataset as presented with Figure 2 -(a).[25] However, in opposite to monthly data, Figure 5 shows a stronger variability in the demand level for a constant temperature, inducing significant effects from other sources than temperature.

Table 8: HDV estimated threshold for daily electricity demand

		$w_{base:1}$	$w_{base:2}$
$q_t^{Daily_e}$	temp	15.27	10.38
$q_t^{Daily_e}$	sunlight	552.72	460.70
$q_t^{Daily_e}$	wind	4.36	4.30
$q_t^{Daily_e}$	rain	5.24	4.19

Notes : The base indicators are defined using K-means algorithm in a two-dimensional environment: energy consumption and level of the indicator. The table shows the base indicator for the regime-switching between the non-heating and heating period. Daily data span 01/01/2019-31/12/2023

Figure 5: Clustering analysis of daily low tension electricity response to temperature



Notes : This figure is a visual representation of the K-means algorithm results for three different regimes. The vertical red line represents the regime-switching point, or base temperature, between the non-heating and heating periods. Daily energy demand span the period 01/01/2019-31/12/2023.

[25]The figures for the other weather variables are available upon request.

Table 9: Daily estimates of low tension electricity response to weather sensibility

	(1)	(2)	(3)	(4)
$HDV_{temp=17}$	0,0139***	(-)	(-)	(-)
	(52.33)			
$HDV_{temp=EU}$	(-)	0,0130***	(-)	(-)
		(50.94)		
$HDV_{temp=15}$	(-)	(-)	0,0149***	0,0153***
			(53.85)	(54.84)
$HDV_{wind=4}$	(-)	(-)	(-)	0,0063***
				(5.33)
$HDV_{sunlight=300}$	(-)	(-)	(-)	0,0001***
				(3.55)
AIC	-6948	-6898	-7003	-7053
RMSE	0.0224	0.0228	0.0220	0.0216

Notes: The estimates are from a regSARIMA regression with daily energy consumption level in TWh as the dependent variable. The table presents the estimates for weather components that significantly impact energy consumption. The estimates for HDV_{temp} variables can be interpreted as the thermo-sensibility to cold temperatures. Then, the estimated coefficients to the HDV_{wind} variables can be interpreted as the wind-sensibility to strong wind speed and the $HDV_{sunlight}$ variables as the sunlight-sensibility to lower sunlight duration. However, the model also estimates the SARIMA components, i.e. the cyclic elements, and control variables for business days: leap year, weekend and holidays effects. Daily energy consumption span the period 01/01/2019-31/12/2023.

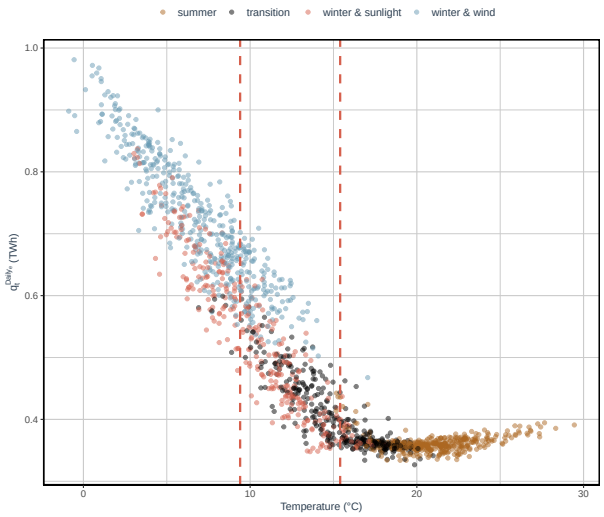
Table 9 confirms that the optimal specification for the regressor variable w_t at the daily level is to introduce a $HDV_{temp=15}$ with a base temperature of 15°C. It leads to the estimation with the lowest AIC and gives an estimate of 0.0153 TWh/$HDV_{temp=15}$, meaning that an increase of one heating degree day, i.e. a decrease in temperature, induces an increase in low voltage electricity demand of 0.0153 TWh. In addition, an increase of one HDV_{wind}, i.e. an increase in wind speed, and one $HDV_{sunligh}$, i.e. a decrease in sunlight duration, increase the low voltage power demand by 0.0065 TWh and 0.0001 TWh, respectively.

Figure 5 shows that the relationship between temperature and low voltage energy demand has a higher variability on a daily frequency than on a monthly frequency, especially for lower temperatures. This implies a variability in demand levels for a constant temperature. To explain this variability, we apply a K-means clustering procedure to a reduced form of the vectors: $g_t^{Daily_e}$; [temp]; [wind]; [sunlight]. The dimension reduction technique condenses information from multiple dimensions into a lower dimensional representation to better capture the essence of high dimensional interactions, which is critical since K-means performance can degrade in higher dimensional spaces.

Various dimension reduction techniques have been developed to preserve different properties of the original data. These techniques range from linear projections, such as Principal Component Analysis (PCA) Wold et al. (1987), to locally linear and nonlinear methods, such as t-Distributed Stochastic Neighbor Embedding (t-SNE) Van Der Maaten and Hinton (2008). In contrast to linear methods, t-SNE focuses on preserving local similarities between data points in high-dimensional space, which facilitates effective representation in a lower-dimensional space. Applying K-means to the reduced t-SNE form of our vectors reveals four distinct clusters: a summer cluster, a transition cluster, and two separate winter clusters. The winter clusters correspond to days with low temperatures and wind, and days with low temperatures and longer sunlight duration. These clusters reflect an increase or decrease in demand relative to the average demand at constant temperature. Thus, to fully account for the demand induced by climate variability at the daily level, it is crucial to include the effects of wind and sunlight. Furthermore,

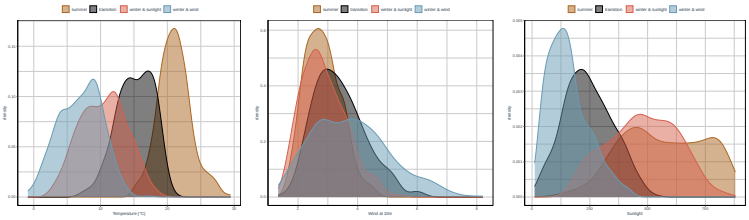
the transition cluster is centered around a temperature of 15°C (Figure 7 -(a)), which is consistent with the overall results of the study.

Figure 6: Clustering analysis of daily low tension electricity demand via tSNE dimension reduction



Notes : This figure is a visual representation of the K-means algorithm results on a four-dimensional dataset reduced to two-dimensions via the t-SNE non-linear method. It allows to put forward the effect of sunlight duration and wind speed during winter compared to temperature and electricity demand. Daily energy demand span the period 01/01/2019-31/12/2023.

Figure 7: Weather variables distribution based on clustering method via tSNE dimension reduction



(a) (b) (c)

Notes : (a)Temperature (b) Wind (c) Sunlight. These figures are a visual representation of the distribution from weather variables, regarding the cluster determined by the K-means algorithm on the reduced form by t-SNE of the daily electricity demand. This figure allow to define the name of each cluster: for example, the cluster "*winter & sunlight*" has a temperature distribution centred around 9°C, a sunlight distribution centred around 500 minutes/24h and very low level of wind and rain. Daily energy demand span the period 01/01/2019-31/12/2023.

5.3 Regional data

This section used a daily disaggregated dataset for each of the 12 regions, to unravel the optimal weather vectors across France. We present the first step using the K-means procedure and it defines an average, normalised, base temperature of 15°C as reported in Table 10. However, the regional decomposition shows a non-continuity in the base temperature detected for each region. In particular, the region *Occitanie*, and *Provence-Alpes-Côte d'Azur* shows a HDV_{temp} and CDV_{temp} related to the temperature effects. Appendix B.1 shows how to optimize the base temperature for CDV_{temp} with our approach when the response function can be described as a U-shaped relationship as is the case for *Provence-Alpes-Côte d'Azur*, meaning that there is an increase in energy demand when temperatures are high for these regions.

Furthermore, Table 10 illustrates the baseline temperature using the clustering algorithm and the SETAR procedure. The results demonstrate that the K-means approach produces a more robust estimate in the presence of spatial heterogeneity. The columns labelled *RMSE* show the root mean square error (RMSE) for the regSARIMA process using the two methods. The results demonstrate that the temperatures extracted by K-means led to a more optimal estimation than with SETAR. This indicates that clustering methods are more robust to spatial heterogeneity. However, both methods produce comparable results at the aggregate level (see appendix B.2 for details).

Table 10: Estimated base temperature for regionals HDV

	$HDV_{temp}^{K-means}$			HDV_{temp}^{SETAR}		
	estimated	round	RMSE	estimated	round	RMSE
Grand Est	13.95	14	0.91	14.46	14	0.91
Normandie	14.10	14	1.79	12.93	13	1.92
Hauts-de-France	14.28	14	1.33	13.27	13	1.41
Bretagne	14.32	14	1.88	15.27	15	1.77
Auvergne-Rhône-Alpes	14.61	15	1.35	17.93	18	1.46
Centre-Val de Loire	14.69	15	1.67	12.86	13	1.88
Bourgogne-Franche-Comté	14.73	15	1.23	16.03	16	1.27
Pays de la Loire	14.91	15	1.79	16.80	17	1.75
Occitanie	15.18	15	1.56	19.14	19	1.50
Île-de-France	15.20	15	1.32	12.73	13	1.51
Nouvelle-Aquitaine	16.00	16	1.49	18.69	19	1.60
Provence-Alpes-Côte d'Azur	16.08	16	1.38	19.20	19	1.42
Mean	14.83	15	1.47	15.86	16	1.53
Weighted Mean	14.94	15	-	16.01	16	-

Notes : The base temperatures are defined using K-means algorithm in a two-dimensional environment: energy consumption and level of temperature. Then, the lowest temperature of each regime is extracted and registered as the base temperature for regime switching. The table shows the base temperature for the regime-switching between the non-heating and heating periods. The national weighted mean is computed using the number of housing recorded on the French national census from 2020. Regional daily energy consumption span the period 01/01/2022-31/12/2023.

Table 11 presents the estimated weather coefficients with a regSARIMA specification. For clarity purposes, only the estimated coefficients are reported and if the estimate was not significant than it has been set to 0. In terms of the level of the estimates, it represents the variation of demand in kWh per household for a variation of one unit from the different weather indicators. Table 11 is shorted according to the thermo-sensitive estimate level, from the lowest to the highest one. Thus, the strongest thermo-sensitivity

coefficient is estimated for the region *Normandie* and the smallest for the region *Grand Est*, highlighting the heterogeneity in the estimated thermo-sensitivity, see Figure 8 for a visual representation.

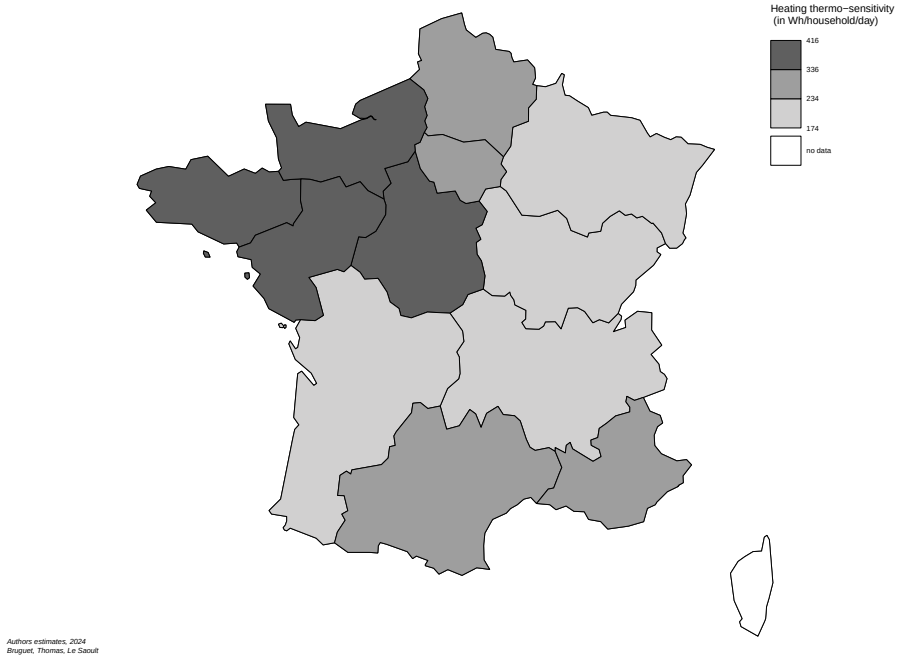
Regarding the GWI_{q_t}, there is heterogeneity depending on the climate zone that the administrative region depends on. First, the correlation table (see Table 20 in appendix A.3) shows that the temperature correlation is strong for all regions, but can be relatively smaller for regions from hot climatic zones, where there is probably a cooling degree effect. Then, the sunlight and wind correlations are on average lower than at the national level. The estimation on Table 11 can be interpreted in terms of GWI_{q_t}, the estimation allows us to develop our interpretation in terms of the pool of significant weather indicators for each region. First, we again identify two regions where there is a significant cooling effect: *Occitanie* and *Provence-Alpes-Côte d'Azur*. These two regions are all located in hot climatic zones in the southern metropolitan area of France. Secondly, the effect of sunlight becomes insignificant at this spatially disaggregated level, and finally, wind speed is significant with a positive estimate, meaning that an increase in wind speed leads to an increase in demand, as at the other time and spatial frequency.

Table 11: Regional daily estimates of low tension electricity response to weather sensibility

	$\boldsymbol{HDV}_{temp}^{K-means}$	$\boldsymbol{CDV}_{temp=21}$	$\boldsymbol{HDV}_{wind=3.5}$	$\boldsymbol{HDV}_{sunlight=400}$	$\boldsymbol{HDV}_{rain=2}$
Grand Est	0.1743	0.0000	0.1317	0.0000	0.0000
Auvergne-Rhône-Alpes	0.1880	0.0000	0.1122	0.0000	0.0000
Bourgogne-Franche-Comté	0.2055	0.0000	0.2075	0.0000	0.0000
Nouvelle-Aquitaine	0.2225	0.0000	0.1724	0.0000	0.0000
Occitanie	0.2464	0.0767	0.1431	0.0000	0.0000
Provence-Alpes-Côte d'Azur	0.2509	0.1533	0.1492	0.0000	0.0000
Île-de-France	0.2705	0.0000	0.2547	0.0000	0.0000
Hauts-de-France	0.3188	0.0000	0.1584	0.0000	0.0000
Bretagne	0.3523	0.0000	0.0000	0.0000	0.0000
Pays de la Loire	0.3717	0.0000	0.2117	0.0000	0.0000
Centre-Val de Loire	0.4018	0.0000	0.3246	0.0000	0.0000
Normandie	0.4159	0.0000	0.2162	0.0000	0.0000

Notes : The estimates are from a regSARIMA regression with regional daily energy demand level in kWh/household as the dependent variable. For comprehension purposes, the non-significant estimates have been reduced to 0. The table presents the estimates for weather components that significantly impact energy demand. The estimates for HDV_{temp} variables can be interpreted as the thermo-sensibility to cold temperatures. Then, the estimated coefficients to the HDV_{wind} variables can be interpreted as the wind-sensibility to strong wind speed, $HDV_{sunlight}$ variables as the sunlight-sensibility to lower sunlight duration and HDV_{rain} as the rain-sensibility to an increase in the millimetres of rain during 24h. However, the model also estimates the SARIMA components, i.e. the cyclic elements, and control variables for business days: leap year, weekend and holidays effects. Regional daily energy demand span the period 01/01/2022-31/12/2023.

Figure 8: Administrative regional daily weather estimates



(a)

(b)

Notes : The maps show the (a) HDV_{temp} estimates and (b) CDV_{temp} estimates from a regSARIMA regression with regional daily energy demand level in Wh/household as the dependent variable. Regional daily energy demand span the period 01/01/2022-31/12/2023.

6 Conclusion

The need for reliable comparisons in energy demand analysis necessitates the adjustment of energy demand for weather variations, allowing for a more accurate isolation of socio-economic factors. Traditional approaches, such as the use of Heating Degree Days (HDD) with a base temperature commonly set at 18°C, have lacked robust methodological justification, particularly in the context of evolving climatic conditions and spatial and sector heterogeneity. Our study challenges these traditional methods by demonstrating that the base temperature is not constant over time and is not uniform across different regions or economic sectors. Furthermore, relying solely on temperature as the indicator for weather variations is insufficient, as other variables such as wind speed and sunlight duration also significantly impact energy demand. To address these limitations, we introduced the General Weather Indicator (GWI), which optimally combines heating days variable (HDV) and cooling days variable (CDV) across multiple weather factors among temperature, wind, sunlight duration, rain and cloudiness. For each variable, we extract the optimal(s) threshold(s) that capture the non-linear relationship between the energy demand and this variable using K-means. Then for each energy demand we select the optimal linear combination between heating and cooling variables using a LASSO penalisation. For each specification, we perform an in-sample adjustment based in regSARIMA modeling and we select the optimal specification according to different criteria likelihood-based (AIC) and forecasting error-based (RMSE). This method provides a robust and adaptable methodology for analyzing energy demand response to weather variations.

Applying this methodology to France's electricity and natural gas demand, we found that the optimal base temperature for HDV is 15°C, a significant deviation from the traditionally used 17°C or 18°C. Moreover, the significance of wind and sunlight, mostly in the residential energy demand, emphasizes the importance of felt temperature rather than measured temperature for households. Then, more broadly, our findings reveal an L-shaped response to temperature, with a heating and a comfort period, at the exception of the tertiary sector and southern regions where air conditioning usage is more prevalent inducing the apparition of a cooling period during summer. The implications of our study are important for energy policy, climate adaptation strategies, and economic planning. By estimating optimaly weather-related elasticities for energy demand at daily and monthly levels across different sectors and regions, we provide a valuable tool for policymakers and researchers. The estimation of these elasticities is of primary importance to calibrate the micro-simulation models which are available for France. This tool enhances the precision of energy consumption analyses in the face of climate variability, supporting the development of more effective strategies to meet emission reduction targets and improve infrastructure planning.

Overall, our study underscores the critical importance of dynamic, region-specific and sector-specific adjustments in energy demand models, thereby improving the accuracy of policy assessments and contributing to the broader goals of climate change mitigation and sustainable energy management.

References

ADEME, 2022. Représentations sociales du changement climatique. Technical Report.

Aghabozorgi, S., Shirkhorshidi, A.S., Wah, T.Y., 2015. Time-series clustering – A decade review. Information systems 53, 16–38. doi:10.1016/j.is.2015.04.007.

Alberini, A., Gans, W., Velez-Lopez, D., 2011. Residential consumption of gas and electricity in the us: The role of prices and income. Energy Economics 33, 870–881.

Asadoorian, M.O., Eckaus, R.S., Schlosser, C.A., 2008. Modeling climate feedbacks to electricity demand: The case of China. Energy economics 30, 1577–1602. doi:10.1016/j.eneco.2007.02.003.

Aubin, C., Fougere, D., Husson, E., Ivaldi, M., 1995. Real-time pricing of electricity for residential customers: Econometric analysis of an experiment. Journal of Applied Econometrics 10, S171–S191.

Bernard, J.T., Idoudi, N., Khalaf, L., Yélou, C., 2007. Finite sample inference methods for dynamic energy demand models. Journal of Applied Econometrics 22, 1211–1226.

Bessec, M., Fouquau, J., 2008. The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach. Energy Economics 30, 2705–2721. doi:10.1016/j.eneco.2008.02.003.

Carcedo, J.M., Vicéns-Otero, J., 2005. Modelling the non-linear response of Spanish electricity demand to temperature variations. Energy economics 27, 477–494. doi:10.1016/j.eneco.2005.01.003.

Chan, K., Petruccelli, J.D., Tong, H., Woolford, S.W., 1985. A multiple-threshold AR(1) model. Journal of Applied Probability/Journal of applied probability 22, 267–279. doi:10.2307/3213771.

Chaton, C., 2024. Impact of public policies on the dynamics of energy retrofit and fuel poverty in mainland France. Working Paper doi:10.2139/ssrn.4743235.

Considine, T.J., 2000. The impacts of weather variations on energy demand and carbon emissions. Resource and energy economics 22, 295–314. doi:10.1016/s0928-7655(00)00027-0.

Cui, X., Gafarov, B., Ghanem, D., Kuffner, T.A., 2023. On model selection criteria for climate change impact studies. Journal of Econometrics , 105511doi:10.1016/j.jeconom.2023.105511.

De Azevedo, J.A., Chapman, L., Muller, C., 2015. Critique and suggested modifications of the degree days methodology to enable long-term electricity consumption assessments: a case study in Birmingham, UK. Meteorological Applications 22, 789–796. doi:10.1002/met.1525.

De Dear, R., Brager, G., 2001. The adaptive model of thermal comfort and energy conservation in the built environment. International journal of biometeorology 45, 100–108. doi:10.1007/s004840100093.

Dell, M., Jones, B.F., Olken, B.A., 2014. What Do We Learn from the Weather? The New Climate-Economy Literature. Journal of Economic Literature 52, 740–798. doi:10.1257/jel.52.3.740.

Donovan, J.J., Fischer, W.P., 1976. Factors affecting residential heating energy consumption. Working Paper .

Dubin, J.A., 2008. An integrated engineering–econometric analysis of residential balance point temperatures. Energy economics 30, 2537–2551. doi:10.1016/j.eneco.2007.02.013.

Engle, R.F., Granger, C.W.J., Rice, J.A., Weiss, A., 1986. Semiparametric Estimates of the Relation between Weather and Electricity Sales. Journal of the American Statistical Association 81, 310–320. doi:10.1080/01621459.1986.10478274.

Eskeland, G.S., Mideksa, T.K., 2010. Electricity demand in a changing climate. Mitigation and adaptation strategies for global change 15, 877–897. doi:10.1007/s11027-010-9246-x.

Fazeli, R., Rüth, M., Davíðsdóttir, B., 2016. Temperature response functions for residential energy demand – A review of models. Urban climate 15, 45–59. doi:10.1016/j.uclim.2016.01.001.

Fowlie, M., Greenstone, M., Wolfram, C., 2018. Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program*. The Quarterly journal of economics 133, 1597–1644. doi:10.1093/qje/qjy005.

Gelegenis, J., 2009. A simplified quadratic expression for the approximate estimation of heating degree-days to any base temperature. Applied energy 86, 1986–1994. doi:10.1016/j.apenergy.2009.02.007.

Gerarden, T., Newell, R.G., Stavins, R.N., 2017. Assessing the Energy-Efficiency Gap. Journal of economic literature 55, 1486–1525. doi:10.1257/jel.20161360.

Giraudet, L.G., Bourgeois, C., Quirion, P., 2021. Policies for low-carbon and affordable home heating: A French outlook. Energy policy 151, 112140. doi:10.1016/j.enpol.2021.112140.

Greevy, J.M., Muguruza, A., Issa, Z., Salvi, C., Chan, J.R., Žurič, Z., 2024. Detecting multivariate market regimes via clustering algorithms. Social Science Research Network doi:10.2139/ssrn.4758243.

Hekkenberg, M., Moll, H., Uiterkamp, A.S., 2009. Dynamic temperature dependence patterns in future energy demand models in the context of climate change. Energy 34, 1797–1806. doi:10.1016/j.energy.2009.07.037.

Henley, A., Peirson, J., 1997. Non-Linearities in Electricity Demand and Temperature: Parametric Versus Non-Parametric Methods. Oxford bulletin of economics and statistics 59, 149–162. doi:10.1111/1468-0084.00054.

Horvath, B., Issa, Z., Muguruza, A., 2021. Clustering Market Regimes Using the Wasserstein Distance. Social Science Research Network doi:10.2139/ssrn.3947905.

Hurn, A.S., Silvennoinen, A., Teräsvirta, T., 2016. A smooth transition logit model of the effects of deregulation in the electricity market. Journal of Applied Econometrics 31, 707–733.

Isaac, M., Van Vuuren, D., 2009. Modeling global residential sector energy demand for heating and air conditioning in the context of climate change. Energy policy 37, 507–521. doi:10.1016/j.enpol.2008.09.051.

Kennard, H., Oreszczyn, T., Mistry, M., Hamilton, I., 2022. Population-weighted degree-days: the global shift between heating and cooling. Energy and Buildings 271, 112315. doi:10.1016/j.enbuild.2022.112315.

Kissock, J.K., Haberl, J., Claridge, D., 2003. Inverse Modeling Toolkit: Numerical Algorithms for Best-Fit Variable-Base Degree Day and Change Point Models. Working Paper .

Koenig, L.R., 1979. Effects of industrial effluents on local cloudiness and rainfall. Water, Air, and Soil Pollution 12, 47–69.

Labriet, M., 2013. Impacts of Climate Change on Heating and Cooling: A Worldwide Estimate from Energy and Macro-Economic Perspective. Working Paper .

Lefieux, V., 2007. Modèles semi-paramétriques appliqués à la prévision des séries temporelles. cas de la consommation d'électricité. working paper .

Liao, T.W., 2005. Clustering of time series data—a survey. Pattern recognition 38, 1857–1874. doi:10.1016/j.patcog.2005.01.025.

Lo, M.C., Zivot, E., 2001. Threshold cointegration and nonlinear adjustment to the law of one price. Macroeconomic Dynamics 5, 533–576. doi:10.1017/s1365100501023057.

Lundström, L., 2017. Adaptive weather correction of energy consumption data. Energy Procedia 105, 3397–3402. doi:10.1016/j.egypro.2017.03.778.

MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. Working Paper 1, 281–297.

Mitchell, J.F.B., 1984. Climate and energy systems: A review of their interactions. Quarterly journal of the Royal Meteorological Society 110, 776–777. doi:10.1002/qj.49711046519.

MTE, 2020. Stratégie nationale Bas-Carbonne. Technical Report.

Pang, O., Bell, W., Monsell, B., 2022. Accomodating Weather Effects in Seasonal Adjustment . Working Paper .

Pardo, Meneu, V., Valor, E., 2002. Temperature and seasonality influences on Spanish electricity load. Energy economics 24, 55–70. doi:10.1016/s0140-9883(01)00082-2.

Peñasco, C., Anadón, L.D., 2023. Assessing the effectiveness of energy efficiency measures in the residential sector gas consumption through dynamic treatment effects: Evidence from England and Wales. Energy economics 117, 106435. doi:10.1016/j.eneco.2022.106435.

Quartier-La-Tente, A., Michalek, A., Palate, J., Baeyens, R., 2024. Interface to 'JDemetra+' Seasonal Adjustment Software [R package RJDemetra version 0.2.5]. R Package .

Rüth, M., Lin, A.C., 2006. Regional energy demand and adaptations to climate change: Methodology and application to the state of Maryland, USA. Energy policy 34, 2820–2833. doi:10.1016/j.enpol.2005.04.016.

Sailor, D.J., Pavlova, A., 2003. Air conditioning market saturation and long-term response of residential cooling energy demand to climate change. Energy 28, 941–951. doi:10.1016/s0360-5442(03)00033-1.

Sgarlato, R., Ziel, F., 2023. The Role of Weather Predictions in Electricity Price Forecasting Beyond the Day-Ahead Horizon. IEEE transactions on power systems 38, 2500–2511. doi:10.1109/tpwrs.2022.3180119.

Sherman, M.H., 1987. Estimation of infiltration from leakage and climate indicators. Energy and Buildings 10, 81–86. doi:10.1016/0378-7788(87)90008-9.

Sinnott, D., 2016. Dwelling Airtightness: a Socio-technical evaluation in an Irish context. Building and Environment 95, 264–271. doi:10.1016/j.buildenv.2015.08.022.

Thao Khamsing, W., Ceci-Renaud, N., Guillot, L., 2016. Simuler l'impact social de la fiscalité énergétique : le modèle Prometheus. Working Paper 138.

Thom, H.C.S., 1954. The rational relationship between heating degree days and temperature. Monthly Weather Review 82.

Van Der Maaten, L., Hinton, G.E., 2008. Visualizing Data using t-SNE. Journal Of Machine Learning Research 9, 2579–2605.

Wold, S., Esbensen, K.H., Geladi, P., 1987. Principal component analysis. Chemometrics and intelligent laboratory systems 2, 37–52. doi:10.1016/0169-7439(87)80084-9.

A Appendix

A.1 Litterature review on HDD/CDD

Table 12: Panel of base temperatures found in the literature and associated justifications

Country	Authors	Journal	Base Temperature (°C)	Justifications
Australia	Badescu and Zamfir (1999)	Energy conversion and management	18	18°C is a common base temperature used in the determination of Heating Degree Days
New Zealand	Badescu and Zamfir (1999)	Energy conversion and management	15,6 and 16	-
Romania	Badescu and Zamfir (1999)	Energy conversion and management	18	18°C is a common base temperature used in the determination of Heating Degree Days
United Kingdom	Badescu and Zamfir (1999)	Energy conversion and management	16	-
Israel	Beenstock et al. (1999)	Energy Economies	10	-
Argentina	Castaneda and Claus (2013)	International journal of climatology	18,3	Thom (1954)
United States of America	Considine (2000)	Resource and Energy Economics	18,3	'A degree-day is the difference between a day's average temperature in Fahrenheit and 65°F'
South-Africa	D. Conradie et al. (2018)	Building Research and Information	18	18°C is a common base temperature used in the determination of Heating Degree Days
Italy	De Rosa et al. (2014)	Applied energy	18,3	Thom (1954)
Saudi Arabia	El-Shaarawi and Al-Masri (1996)	Energy	17,8 and 21,1	-
Europe	Eskeland and Mideksa (2010)	Mitigation and adaptation strategies for global change	18 and 22	'A temperature interval is defined as a comfort zone [. . .] i.e., between 18 and 22°C"
Western Europe	Golombek et al. (2012)	Climatic change	18	18°C is a common base temperature used in the determination of Heating Degree Days
Netherlands	Hekkenberg et al. (2009)	Energy Policy	18	18°C is a common base temperature used in the determination of Heating Degree Days
Wordlwide	Isaac and Van Vuuren (2009)	Energy Policy	18	18°C is a common base temperature used in the determination of Heating Degree Days
Turkey	Kadioglu et al. (1999)	Applied Meteorology and Climatology	15	18°C is a common base temperature used in the determination of Heating Degree Days
Hong Kong	Lam (1998)	Energy conversion and management	18,3	Thom (1954)
Greece	Papakostas at al. (2010)	Renewable Energy	15	-
Saudi Arabia	Said (1992)	Engineering And Applied Engineering	18 and 21	-
Ireland	Semmler et al. (2010)	Meterological Applications	18	"This temperatures is the most common base temperature of normally insulated buildings'
Europe	Spinoni et al. (2015)	International journal of climatology	15,5	-
United States of America	Taha (1997)	Energy and buildings	18,3	Thom (1954)
Macedonia	Taseska et al. (2012)	Energy	20	-
United States of America	Thom (1954)	Monthly Weather Review	18,3	Derived from a historical temperature probability function
Spain	Valor et al. (2001)	Journal of Applied Meteorology	15	'Within these two limits a comfort zone was established and no heating or cooling is required'
Spain	Valor et al. (2001)	Journal of Applied Meteorology	18	18°C is a common base temperature used in the determination of Heating Degree Days
United States of America	Deschênes and Greenstone (2011)	American economic journal	18,3	Thom (1954)

A.2 Data sources

Table 13: Variables used, sources and transformations

Data category	Variable	Code	Period	Data Source	Original Code	Scale
Energy consumption	Electricity consumption at low tension	conso_elec	01/2000 - 10/2023	French Statistical Data and Studies Department	CONSO_ELEC_BT	TWh
	Electricity consumption at low tension	conso_elec	01/2000 - 10/2023	French Statistical Data and Studies Department	CONSO_ELEC_MT	TWh
	Electricity consumption at low tension	conso_elec	01/2000 - 10/2023	French Statistical Data and Studies Department	CONSO_ELEC_HT	TWh
	Gas distributed consumption	conso_gas	01/2000 - 10/2023	French Statistical Data and Studies Department	CONSO_GAZ_TRANSP	TWh
	Gas transported consumption	conso_gas	01/2000 - 10/2023	French Statistical Data and Studies Department	CONSO_GAZ_DISTRIB	TWh
Energy production	Total net electricity generation	prod_elec	01/2000 - 10/2023	French Statistical Data and Studies Department	PRODN_ELEC	TWh
	Primary production of natural gas	prod_gas	01/2000 - 10/2023	French Statistical Data and Studies Department	PRODRES_GAZ	TWh PCS
	Natural gas import balance	prod_gas	01/2000 - 10/2023	French Statistical Data and Studies Department	QSOLI_GAZ	TWh PCS
Energy price	Spot electricity price	price_elec	01/2000 - 10/2023	French Statistical Data and Studies Department	PRIXG_ELE	€
	Spot gas price (PEG Nord)	price_gas	01/2000 - 10/2023	French Statistical Data and Studies Department	PRIXG_GAZ_PEG	€
	Consumer price index (CPI)	price	01/2000 - 10/2023	National Institute of Statistics and Economic Studies (Insee)	IPC	€
Weather	Cloudiness	cloudiness	01/2000 - 10/2023	French meteorology and climatology service	NBSIGMA80	Days
	Rainfall	rain	01/2000 - 10/2023	French meteorology and climatology service	RR	Millimeters
	Sunlight	sunlight	01/2000 - 10/2023	French meteorology and climatology service	INST	Minutes
	Temperature	temp	01/2000 - 10/2023	French meteorology and climatology service	TNTXM	°C
	Wind speed	wind	01/2000 - 10/2023	French meteorology and climatology service	FFM	Meters per second
Demography	France total population	pop	2020	National Institute of Statistics and Economic Studies (Insee)	PTOT	Number of individual
Working days	Working days - Week	reg2_ac1	01/2000 - 10/2023	National Institute of Statistics and Economic Studies (Insee)	REG2_AC1	
	Working days - Week-end	reg2_ac2	01/2000 - 10/2023	National Institute of Statistics and Economic Studies (Insee)	REG2_AC2	
	Leap year	ly	01/2000 - 10/2023	National Institute of Statistics and Economic Studies (Insee)	LY	

A.3 Descriptive statistics

A.3.1 Monthly data

Figure 9: ACF and PACF for Electricity and Natural gas demand

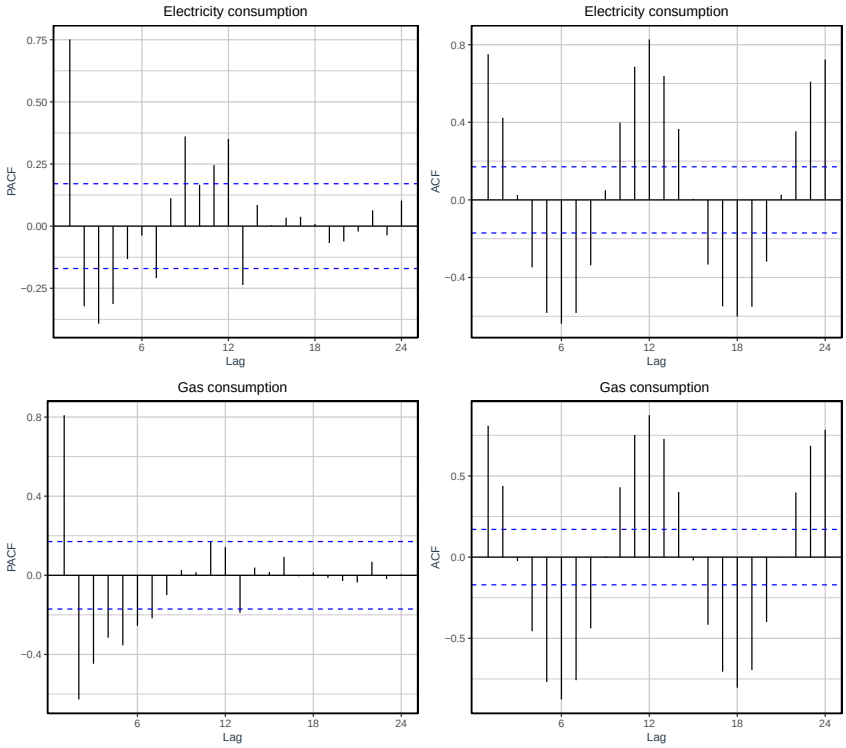


Table 14: Descriptive statistics of demand (TWh)

	N	Mean	St. Dev.	Min	Max
$q_t^{Low_g}$	288	25.71	17.05	5.33	64.16
$q_t^{High_g}$	288	14.11	3.41	8.12	22.41
$q_t^{Low_g}$	288	15.53	4.93	8.89	27.86
$q_t^{Med_g}$	288	12.66	1.73	7.08	16.27
$q_t^{High_g}$	288	7.59	1.28	4.73	9.96
q_t^{gas}	288	39.82	20.02	14.34	84.17
q_t^{elec}	288	35.78	5.85	27.65	51.59

Notes : The descriptive statistics are reported as TWh consumption. The variables q_t^{gas} and q_t^{elec} represent the sums of the respective variables $q_t^{Low_g}$, $q_t^{High_g}$ and $q_t^{Low_g}$,$q_t^{Med_g}$,$q_t^{High_g}$. The monthly data span the period 01/2012-12/2022.

Table 15: Share of economic sector regarding the energy delivery mode - Median

	Residential	**Service**	**Industrial**	**Other**	**Total**
$q_t^{Low_e}$	79,6%	16,0%	1,5%	2,9%	100%
$q_t^{Med_e}$	0,0%	57,5%	39,0%	3,3%	100%
$q_t^{High_e}$	0,0%	18,8%	81,2%	0,0%	100%
$q_t^{Low_g}$	46,4%	26,2%	27,2%	1,3%	101%
$q_t^{High_g}$	0,0%	6,2%	93,8%	0,0%	100%

Notes : The table presents the breakdown of energy consumption by delivery mode into the sectorial consumption mode. This distribution rely on a SDES annual survey of energy suppliers and is computed on the 2018-2022 vintage.

Table 16: Share of delivery mode for each energy demand - Median

	$q_t^{Low_e}$	$q_t^{Med_e}$	$q_t^{High_e}$
q_t^{Elec}	43.44%	35.23%	21.33%
	$q_t^{Low_g}$		$q_t^{High_g}$
q_t^{Gas}	64.62%		35.38%

Notes : The table presents the breakdown of aggregate energy demand time series regarding the delivery mode.

A.3.2 Daily data

Figure 10: Daily electricity demand in TWh

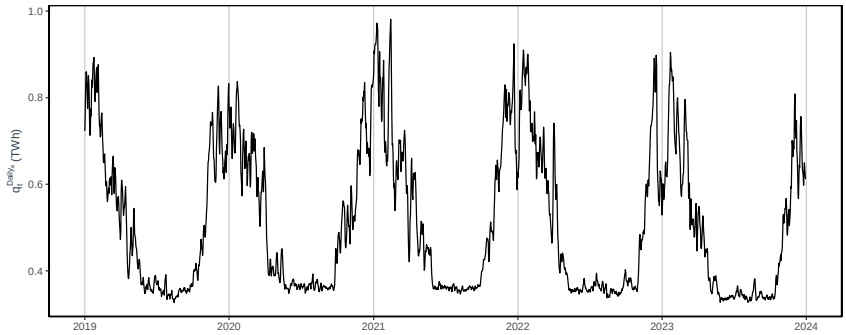


Table 17: Descriptive statistics electricity demand (< 36 kVA) in TWh

	N	Mean	St. Dev.	Min	Max
$q_t^{Daily_e}$	1096	0.524	0.168	0.327	0.981
$q_t^{Monthly_e}$	36	15.95	4.87	10.64	26.22
q_t^{Low}	36	15.92	4.88	10.59	26.04

Notes : The table presents the data distribution at the daily level $q_t^{Daily_e}$ for the 1096 days available on the dataset. The same statistics are computed at the monthly aggregate level $q_t^{Monthly_e}$ on the series and compared with the $q_t^{Low_e}$ monthly series from the main study. Both series from different sources and on different time frames describe similar energy consumption behaviour.

Table 18: Correlation between weather indicators and energy time series - Daily data

	temp	sunlight	wind	rain
$q_t^{Daily_e}$	-0.912	-0.548	0.221	0.042
	0.000	0.000	0.000	0.162

Notes : The table show the correlation coefficients and the p-value associated. The p-values represent the probability that the null hypothesis, which represents a null correlation, is non-rejected. Thus a null p-value is interpreted as a correlation significantly different from 0.

A.3.3 Daily regional data

Figure 11: France 12 administrative regions

Table 19: Descriptive statistics regional electricity demand (< 36 kVA) in TWh

	N	Mean	St. Dev.	Min	Max
Centre-Val de Loire	730	0.021	0.008	0.013	0.041
Bourgogne-Franche-Comté	730	0.022	0.007	0.015	0.042
Normandie	730	0.028	0.011	0.017	0.057
Bretagne	730	0.029	0.011	0.018	0.058
Grand Est	730	0.031	0.010	0.020	0.056
Pays de la Loire	730	0.032	0.012	0.020	0.063
Hauts-de-France	730	0.039	0.013	0.024	0.074
Provence-Alpes-Côte d'Azur	730	0.048	0.014	0.032	0.086
Nouvelle-Aquitaine	730	0.050	0.017	0.033	0.096
Occitanie	730	0.052	0.017	0.035	0.100
Auvergne-Rhône-Alpes	730	0.063	0.021	0.042	0.122
Île-de-France	730	0.074	0.025	0.043	0.138

Notes : The table presents the data distribution at the daily level for 12 different regions and for the 730 days available on the dataset. The region "*Centre-Val de Loire*" is the one with the lowest mean consumption over a day, with 0.021 TWh and the region "*Ile-de-France*" is the one with the highest mean consumption over a day with 0.074TWh.

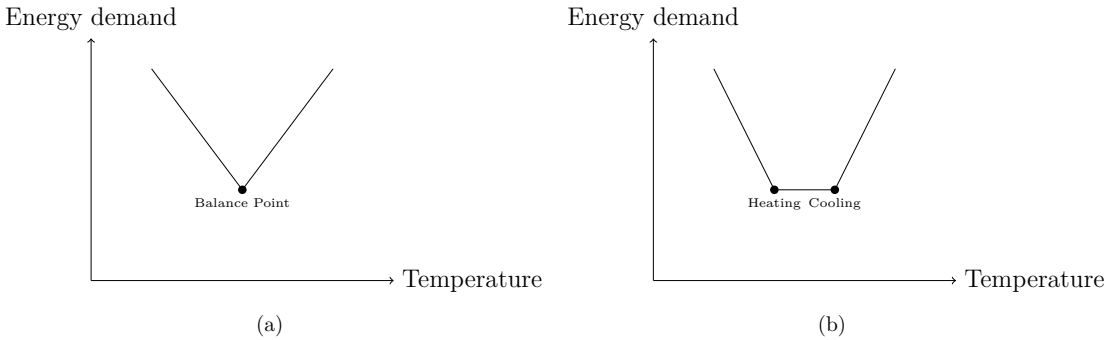
Table 20: Correlation between weather indicators and energy time series - Daily regional data

	temp	sunlight	wind	rain
Île-de-France	-0.906	-0.493	0.168	-0.010
	(0.000)	(0.000)	(0.001)	(0.848)
Centre-Val de Loire	-0.897	-0.522	0.181	-0.097
	(0.000)	(0.000)	(0.001)	(0.065)
Bourgogne-Franche-Comté	-0.904	-0.494	0.182	-0.085
	(0.000)	(0.000)	(0.000)	(0.105)
Normandie	-0.891	-0.484	0.205	0.038
	(0.000)	(0.000)	(0.000)	(0.467)
Hauts-de-France	-0.899	-0.494	0.183	0.015
	(0.000)	(0.000)	(0.000)	(0.780)
Grand Est	-0.911	-0.529	0.214	-0.045
	(0.000)	(0.000)	(0.000)	(0.393)
Pays de la Loire	-0.875	-0.500	0.119	0.016
	(0.000)	(0.000)	(0.023)	(0.767)
Bretagne	-0.870	-0.493	0.111	0.101
	(0.000)	(0.000)	(0.034)	(0.055)
Nouvelle-Aquitaine	-0.866	-0.391	0.056	0.011
	(0.000)	(0.000)	(0.285)	(0.839)
Occitanie	-0.826	-0.345	0.063	0.047
	(0.000)	(0.000)	(0.233)	(0.367)
Auvergne-Rhône-Alpes	-0.888	-0.484	0.018	-0.087
	(0.000)	(0.000)	(0.732)	(0.098)
Provence-Alpes-Côte d'Azur	-0.784	-0.329	0.145	0.002
	(0.000)	(0.000)	(0.005)	(0.970)

Notes : The table show the correlation coefficients and the p-value associated. The p-values represent the probability that the null hypothesis, which represents a null correlation, is non-rejected. Thus a null p-value is interpreted as a correlation significantly different from 0.

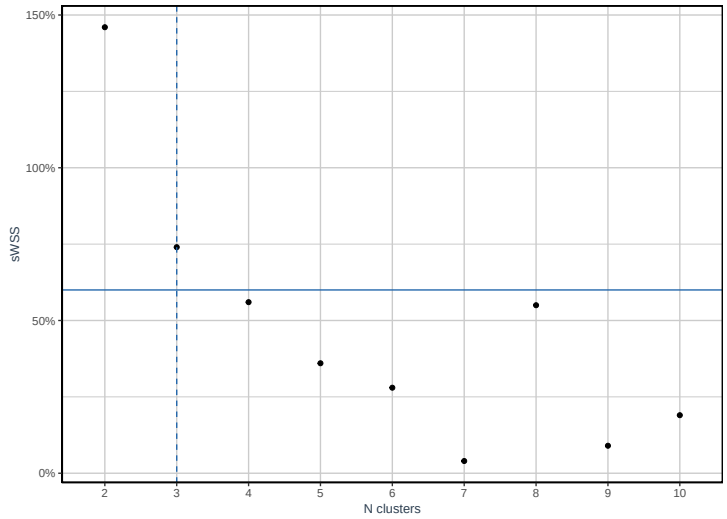
A.4 Functional form for response to temperature

Figure 12: (a) V-shaped relationship - (b) U-shaped relationship



A.5 Procedure to define the number of centroïds

Figure 13: Inertia variation between cluster N and $N-1$ for $q_t^{Low_e}$ versus temperature

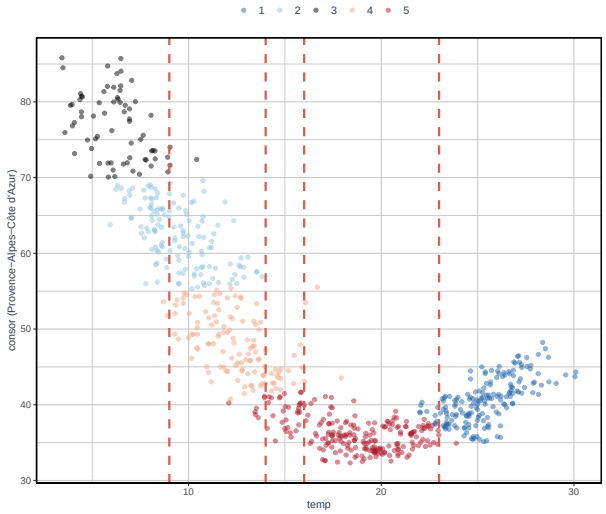


B Complementary results

B.1 Base temperature for Cooling Degree Days

This appendix presents the classification pattern to define using the K-means methodology of the study to be able to take into consideration the energy demand pattern due to high temperature and cooling behaviour, when the response function to temperature can de describe by a U-shaped pattern. It means that the response function has at leas three broad regimes namely : heating and cooling regimes and a so-called comfort zone where the response function is constant. Figure 14 showcases the demand for electricity demand on the low tension line for the *"Provence-Alpes Côte d'Azur"* region of France. It depicts a five regimes adaptation, the base temperature is still 16.1°C for the HDD and for the CDD the method estimates a base temperature of 22.9°C meaning that for days when the average temperature is greater than 22.9°C in this particular region then there is a cooling behaviour and the CDD will be different from zero.[26]

Figure 14: *Provence-Alpes-Côte d'Azur* low tension Electricity



Notes : This figure is a visual representation of the K-means algorithm results for five different regimes. The vertical red line represents the regime-switching point, or base temperature. Daily energy demand span the period 2022-2023 for *Provence-Alpes-Côte d'Azur* region.

[26]For reference, between 2022 and 2023 in Provence-Alpes-Côte d'Azur both months of July had every day an average temperature greater than 22.9°C.

B.2 Threshold detection via SETAR procedure

An initial methodology was to draw upon Carcedo and Vicéns-Otero (2005) and Bessec and Fouquau (2008), which employs a Smooth Transition AR (STAR) method based on a predetermined definition of the data generation function. However, a smooth method induces to define a priori the function that best describes the relation between the series. Since we want our method to be also useful for other indicators than the temperature, our proposal relies on the adoption of a non-smooth methodology grounded in the method of multivariate Threshold Vector Autoregression (TVAR) from Lo and Zivot (2001) and based on the Self-extracting AR (SETAR) modelling paradigm.

A TVAR model extends the traditional VAR framework by incorporating threshold effects, allowing for nonlinear dynamics in the relationships between variables. This approach is particularly useful when the relationships between variables are subject to structural changes or regime shifts. In a multivariate TVAR model, the data is divided into different regimes or states, and separate VAR models are estimated for each regime. The key feature of a multivariate TVAR model is the identification of thresholds that determine the switch between different regimes. For reference, in a univariate dimension, the SETAR modelling with two regimes and one threshold can be describe as in (12) and allows to estimate the coefficients ϕ_1 and ϕ_2 but also the threshold c that correspond to the transition between both regimes.

$$z_t = \begin{cases} \phi_1 z_{t-1} + \epsilon_{1t} & \text{if } z_{t-1} \leq c \\ \phi_2 z_{t-1} + \epsilon_{2t} & \text{if } z_{t-1} > c \end{cases} \tag{12}$$

Estimating the threshold parameter is not obvious due to its representation as a discontinuous function. A viable approach involves concentrating the objective function since the slope estimators given a known threshold can be estimated by ordinary least squares (OLS), the problem can be simplified by concentrating out the minimization problem through $\phi(\Theta)$ and the corresponding sum of squares SSR(Θ). This leads to the following objective function:

$$\hat{\Theta} = \arg\min_{\Theta} SSR(\Theta) \tag{13}$$

Minimization of (13) is done through a grid search: values of the threshold are sorted, the SSR is estimated for each selected threshold and the one that minimize the SSR is taken as the estimator.

Tables 21 showcases thresholds levels estimated with this first method.

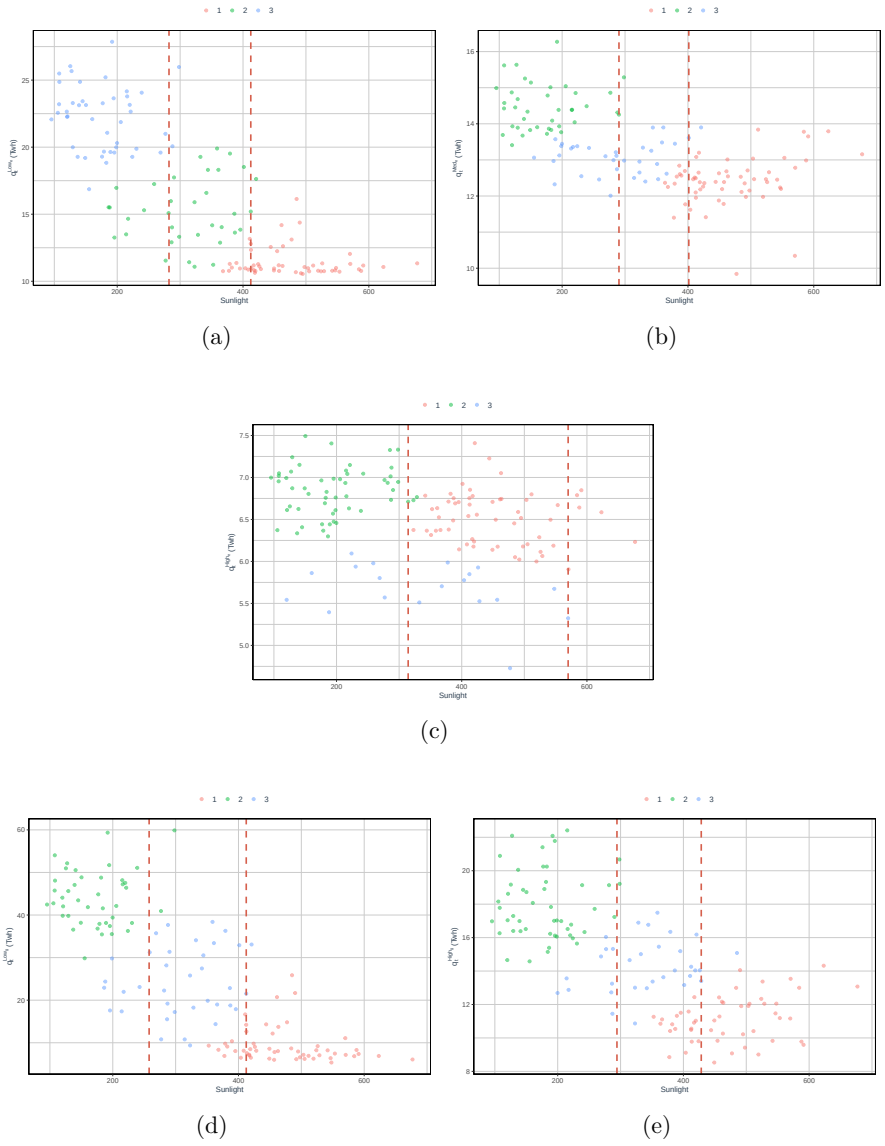
Table 21: Estimated HDV_{base} for monthly energy demands based on SETAR

	Temperature (a)					
	$q_t^{Low_g}$	$q_t^{High_g}$	$q_t^{Low_e}$	$q_t^{Med_e}$	$q_t^{High_e}$	$\hat{w_{base}}$
$w_{base:temp1}$	10.79	9.87	10.02	13.00	13.00	10
$w_{base:temp2}$	15.15	15.50	15.15	18.63	15.79	15
	Sunlight, Wind, Rain and Cloudiness (b)					
	$q_t^{Low_g}$	$q_t^{High_g}$	$q_t^{Low_e}$	$q_t^{Med_e}$	$q_t^{High_e}$	$\hat{w_{base}}$
$w_{base:wind1}$	3.20	3.16	3.16	3.14	3.14	3
$w_{base:wind2}$	3.63	3.53	3.52	3.58	3.49	3.5
$w_{base:sunlight1}$	281	286	323	379	286	300
$w_{base:sunlight2}$	395	395	412	461	395	400
$w_{base:rain1}$	1.58	1.58	1.58	1.58	1.42	1.5
$w_{base:rain2}$	2.21	2.38	2.21	2.34	2.34	2
$w_{base:cloud1}$	25.14	23.33	24.42	24.42	24.42	24

Notes : Extracted thresholds w_{base} using the SETAR procedure. Monthly energy consumptions and weather data span 01/2012-12/2022.

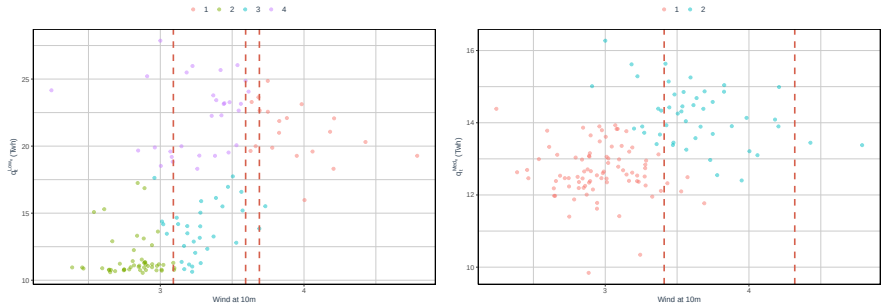
B.3 Threshold detection via K-means beyond temperature

Figure 15: Clustering analysis of monthly energy response to sunlight



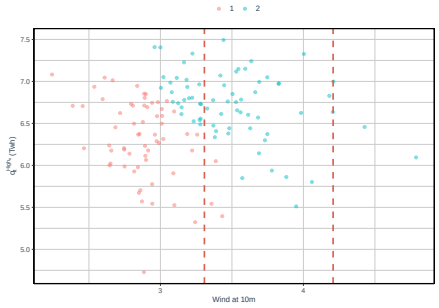
Notes : (a) $q_t^{Low_e}$ (b) $q_t^{Med_e}$ (c) $q_t^{High_e}$ (d) $q_t^{Low_g}$ (e) $q_t^{High_g}$ - These figures are a visual representation of the K-means algorithm results. The dashed lines represent the regime-switching point the clusters represent the different regimes.

Figure 16: Clustering analysis of monthly energy response to wind

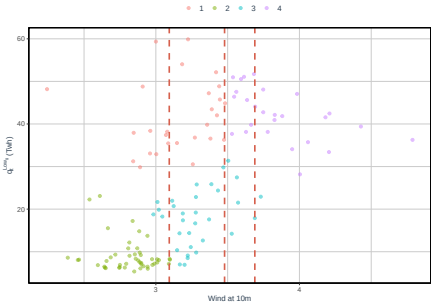


(a)

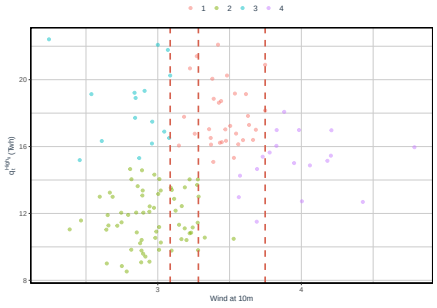
(b)



(c)



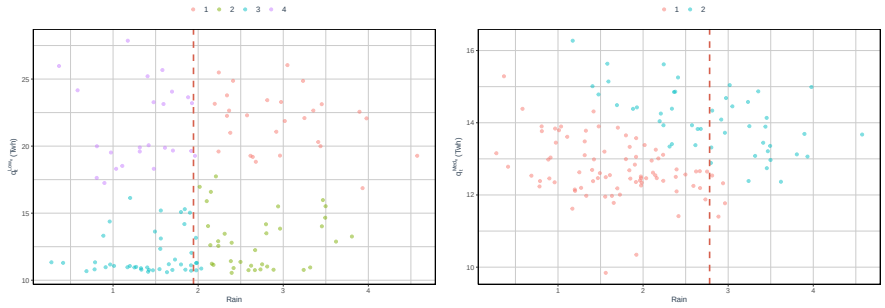
(d)



(e)

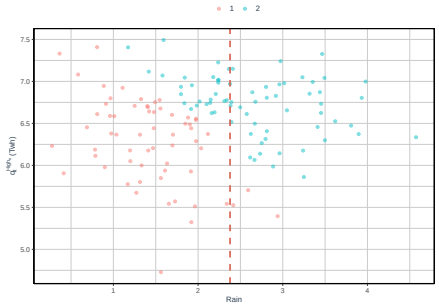
Notes : (a) $q_t^{Low_e}$ (b) $q_t^{Med_e}$ (c) $q_t^{High_e}$ (d) $q_t^{Low_g}$ (e) $q_t^{High_g}$ - These figures are a visual representation of the K-means algorithm results. The dashed lines represent the regime-switching point and the clusters represent the different regimes.

Figure 17: Clustering analysis of monthly energy response to rain

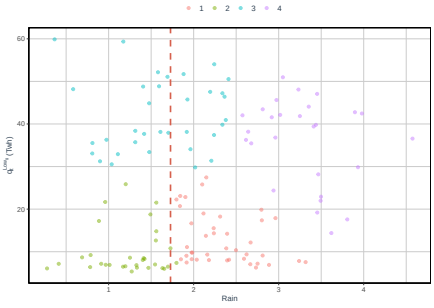


(a)

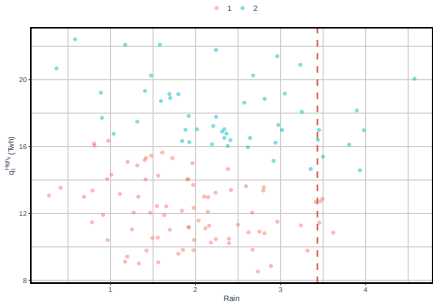
(b)



(c)



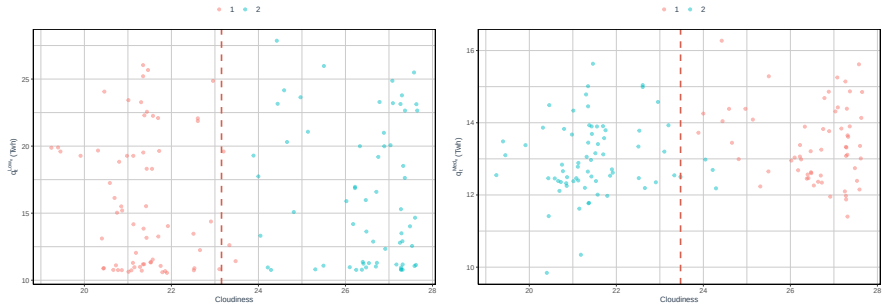
(d)



(e)

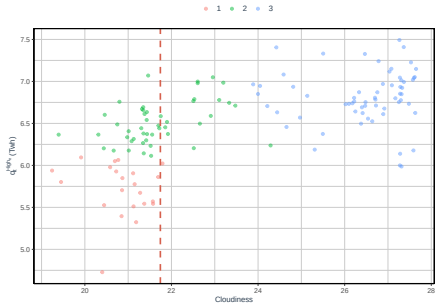
Notes : (a) $q_t^{Low_e}$ (b) $q_t^{Med_e}$ (c) $q_t^{High_e}$ (d) $q_t^{Low_g}$ (e) $q_t^{High_g}$ - These figures are a visual representation of the K-means algorithm results. The dashed lines represent the regime-switching point and the clusters represent the different regimes.

Figure 18: Clustering analysis of monthly energy response to cloudiness

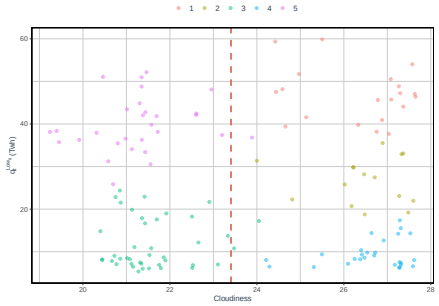


(a)

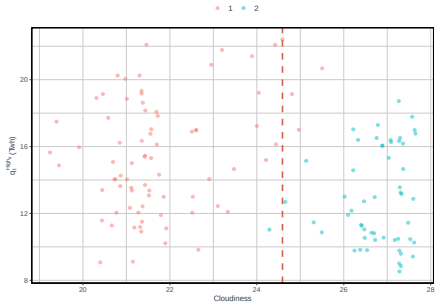
(b)



(c)



(d)

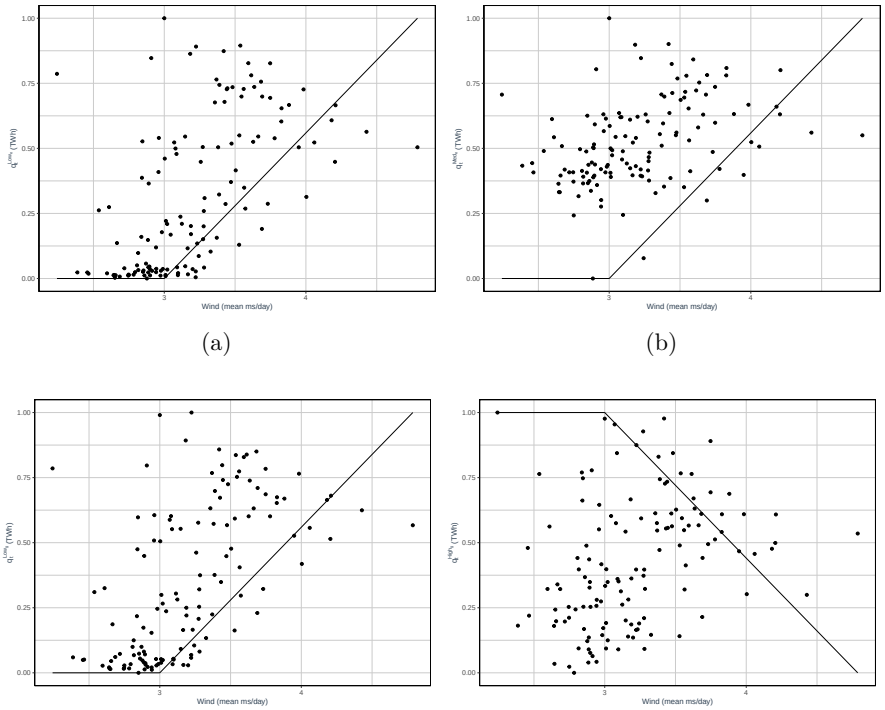


(e)

Notes : (a) $q_t^{Low_e}$ (b) $q_t^{Med_e}$ (c) $q_t^{High_e}$ (d) $q_t^{Low_g}$ (e) $q_t^{High_g}$ - These figures are a visual representation of the K-means algorithm results. The dashed lines represent the regime-switching point and the clusters represent the different regimes.

B.4 Functional form beyond temperature

Figure 19: Linear normalized optimal estimation response to wind



(a) (b)

(c) (d)

Notes :(a) $q_t^{Low_e}$ (b) $q_t^{Med_e}$ (c) $q_t^{Low_g}$ (d) $q_t^{High_g}$ - These figures are a normalised visual representation of the linear estimations of wind sensitivity for the different energy demand. This representation only allows to take into consideration the shape of the relationship but does not correspond to the exact fitted/estimated data on the regression process. Monthly energy demands and weather data span 01/2012-12/2022.