



**HAL**  
open science

## Ex-post energy savings assessment methodologies using smart-meter data: the case study of a switch from direct electric heating to an air-to-air heat pump

Dominique Osso, Guillaume Binet, Benjamin Petiau, Marie-Hélène Laurent

### ► To cite this version:

Dominique Osso, Guillaume Binet, Benjamin Petiau, Marie-Hélène Laurent. Ex-post energy savings assessment methodologies using smart-meter data: the case study of a switch from direct electric heating to an air-to-air heat pump. ECEEE Summer Study Proceedings, European Council for an Energy Efficient Economy, Jun 2024, Lac d'Ailette, France. pp.445. hal-04618563

**HAL Id: hal-04618563**

**<https://edf.hal.science/hal-04618563v1>**

Submitted on 20 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Ex-post energy savings assessment methodologies using smart-meter data: the case study of a switch from direct electric heating to an air-to-air heat pump

Dominique Osso

EDF Lab Les Renardières Department of Technology & Research in Energy Efficiency Avenue des Renardières  
77250 Moret Loing et Orvanne France

G.Binet

EDF Lab Les Renardières Department of Technology & Research in Energy Efficiency Avenue des Renardières  
77250 Moret Loing et Orvanne France

B.Petiau

EDF Lab Les Renardières Department of Technology & Research in Energy Efficiency Avenue des Renardières  
77250 Moret Loing et Orvanne France

M-H.Laurent

EDF Lab Les Renardières Department of Technology & Research in Energy Efficiency Avenue des Renardières  
77250 Moret Loing et Orvanne France

## Abstract

The need to electrify end uses to achieve carbon neutrality by 2050 and strengthen energy independence means that we need to seek out and evaluate the most efficient technologies. To limit these future increases in electricity consumption, existing consumptions must also be reduced. Thus, retrofitting direct electric space heaters with a heat pump is a dual opportunity to speed up the decarbonization of the French residential sector and to moderate both individual and overall electricity demand increase. We then use this case study to evaluate a range of evaluation methods.

To assess the energy savings achieved by the switch from direct electric heating to air-to-air heat pump, in the absence of information to set up a control group (difference in difference), we monitored the daily electricity consumption of a dwelling, via the existing smart meter during 2 periods: 3 years with direct electric heating (DE) before retrofit, and 2 years after the installation of the air-to-air heat pump (HP).

The daily meteorological data were used to assess the thermal gradient (consumption versus temperature). The aim is to capture the heat-sensitive part of the daily consumption (mainly linked to space heating).

Several methods have been used to estimate energy savings by comparing the before and after situations: Linear, quantile and partial least square regressions, Multivariate adaptive regression spline model, and Machine learning based on generalized additive model.

The impact of the energy savings estimated according to the methodology is assessed and the uncertainties that may exist within each method and between methods is presented. The conditions of use and advantages of each method are discussed as they apply to other smart meter consumption, such as gas and other renovation.

## Introduction

The context of a strong end-uses electrification in (building and transportation) to reach carbon neutrality in 2050 (RTE 2024) and to enhance energy independence, leads us to seek after the most efficient technologies to minimize energy consumptions. Thus, the European Commission with the REPowerEU plan, “doubled down its ambition level by communicating the need to increase annual Heat Pumps (HPs) sales by a factor of 2 over the next 5 years” (EHPA, 2022).

Already, the French market for air-to-water HPs has grown from 96,000 units sold per year to 306,000 units between 2018 and 2023. Over the same period, the market for air-to-air heat pumps rose from 573,000 units to

910,000 units<sup>1</sup> sold per year (Uniclimate 2024). France is now the first market in Europe for heat pumps (IEA 2022). France also has many homes heated by Direct Electric (DE) heating (9 million over 30 million) (CEREN 2023) which could be replaced by air-to-air HPs.

To moderate for the future national increase in electricity consumption (Thomaßen, et al. 2021), existing consumption must also be reduced through insulation and efficient equipment such as HPs. In this context, we study the retrofit of a DE-heated dwelling through the implementation of an air-to-air HP after renovation. We analyse the overall electricity consumption (daily) thanks to a smart meter over the last 5 years (2018-2023). This corresponds to 3 winter seasons with DE space heating and 2 winter seasons with the HP.

Beyond the results of this case study and thanks to the smart meter deployment (Wang et al. 2019), our aim is to study generalisable and automatable methodology for a range of situations to assess the modification of consumption (mainly electricity or gas from smart meters).

## CASE STUDY

We monitored the daily electricity consumption of the dwelling; via the smart meter Linky (Enedis 2022) which measures the total electricity delivered to the dwelling, all end-uses combined and a sub-meter for the recent use of a Plug-in Hybrid Electric Vehicle (PHEV). It should be noted that in the following text, we refer for simplification to HP or DE consumptions as the consumption for all end-uses with the heat pump or direct electric as the heating equipment used, unless otherwise specified.

### *The case study*

The studied dwelling is a Single-Family Housing (SFH) built in the end of the '80s located in the "Ile de France" French region (H1a French climatic zone). The typology of this SFH is described by (CSTB, 2022). This house, with a living area of about 130 m<sup>2</sup>, was built in the first wave (around 1982-1995) of the development of DE space heating in France (Osso and Laurent 2017). Subsequently, renovation works were carried out (roof insulation, change of windows) before 2018 (Figure 1).

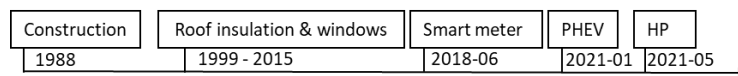


Figure 1. Chronology of events in the case study.

Concerning the end-uses, the SFH is a dwelling fully equipped with electricity end-uses (domestic hot water with storage, cooking) and well equipped with household appliances (washing machine, fridge, dishwasher, tumble dryer, freezer, computer...) and since January 2021 the use of a PHEV with a 12.4 kWh battery (3.7 kW on-board charger). The PHEV is plugged into a standard 16 A power supply each day.

No other notable changes have taken place and we can assume by default that usage behaviour has not drastically changed either. The room-by-room management of space heating by the occupant has not drastically changed, with the heating switched off in summer (and the HP switched to air-conditioning mode in the event of a heatwave). However, as the indoor temperature was not measured, it is impossible to assess the potential rebound effect that would reduce energy savings.

In the case of a large-scale analysis of daily consumption, and in the absence of a detailed household survey, it will be difficult to separate out all these effects. Nevertheless, it is still possible to measure actual energy savings including any undeclared changes, but it is not possible to estimate the performance of equipment. The question that then arises is whether changes in occupancy or behaviour are systematically associated with retrofit or whether this remains marginal.

### *Space heating equipment*

The SFH was initially heated using DE heaters. A multi-split air-to-air HP equipment was installed in May 2021, but electric towel warmers remain in the bathrooms. According to the manufacturer's brochure, the 2 installed external units have a SCOP<sup>2</sup> of 4.7, and a SEER<sup>2</sup> of 7.9 with thermal power from 3 to 9.2 kW (8 kW in cooling mode). The external unit has a nominal electric power of 2.03 kW in heating mode (min=0.7 kW, max=3.06 kW) and 1.98 kW in cooling mode. The HP replaces a total DE installed power of 12 kW.

### *Electricity consumption*

<sup>1</sup> Number of external unit. Multi-splits represent 32% of the total.

<sup>2</sup> SCOP: Seasonal Coefficient of Performance, SEER: Seasonal Energy Efficiency Ratio.

The daily electricity consumption (all end-uses) is measured thanks to the smart meter from June 2018 to August 2023 (Figure 2). It should be noted that in this study we are using real climate consumption data (without climate correction).

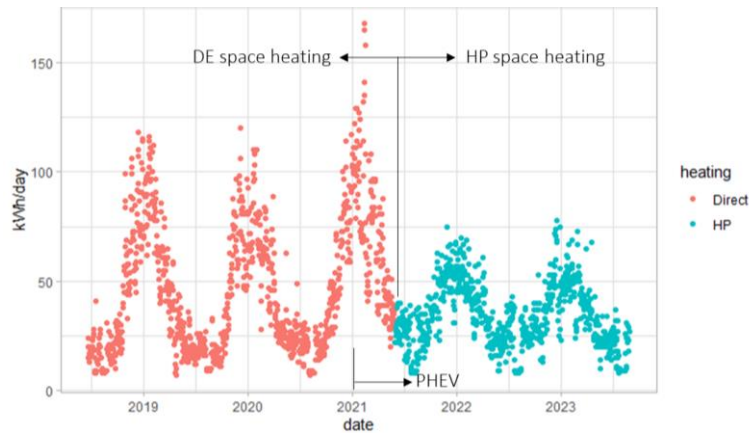


Figure 2. Electric daily consumption (kWh), all end-uses. DE: direct electric, HP: heat pump.

In addition, the load of the PHEV has been monitored via a dedicated sub-meter since August 2022. The measured data are the active instantaneous powers at 6-second intervals. They are processed by Python scripts (Binet et al. 2021). For information, the consumption of the PHEV measured over 12 months (between August 2022 and August 2023) is 2,290 kWh which includes vehicle preconditioning during the cold season. This consumption is included in the daily consumption from the smart meter since January 2021.

### Meteorological data

We used the nearest weather station to the SFH (Meteostat 2022) as a reference, which is located around 20 km from the site. The relative humidity (RH) is provided by (Historique Météo 2023) but for another location (15 km from the site) as (Meteostat 2022) is not providing RH. The lowest average daily temperature ( $T_{average}^d$ ) is  $-3.8^{\circ}\text{C}$  and the highest  $31.9^{\circ}\text{C}$  and an average temperature of  $12.6^{\circ}\text{C}$  (Table 1, grey row). The Figure 3 presents the distribution of explanatory variable values at our disposal.

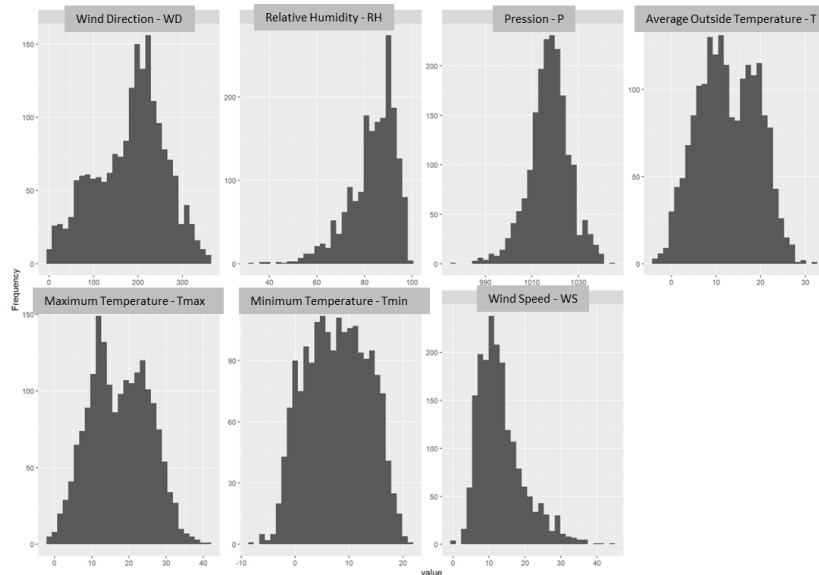


Figure 3. Distribution of explanatory variable values (from 2018-06-21 to 2023-08-28).

Table 1. Observed variables and sources.

Variable	min	max	median	mean	Unit	Source	Code
Date	2018-06-21	2023-08-28	-	-	Day		d
Daily consumption (all end-uses)	7.00	168.00	33.00	40.71	kWh	Smart meter	C
Average outside temperature	-3.85	31.93	12.20	12.67	$^{\circ}\text{C}$	MeteoStat	T
Relative humidity	33.00	100.00	85.00	83.44	%	Historique Météo	RH
Wind speed	0.00	44.32	12.08	13.29	km/h	MeteoStat	WS

Wind direction	0.00	359.0	197.1	184.1	degree	MeteoStat	WD
Atmospheric pressure	976.5	1045.2	1017.7	1017.4	hPa	MeteoStat	P

## Methodology

In this study, our methodology is based on 3 stages:

- Isolation of the days when space heating is active (non-heating temperature) (§ tipping points).
- Modelling of the reference situation (i.e. DE heating consumption) using different statistical models (§ modelling the reference situation).
- Estimation of energy savings by comparing the modelled situation with observed consumption (§energy savings assessment).

Other possibilities were considered for estimating energy savings. A simple comparison at the same average outside temperature between the 2 situations and the creation of 2 models (DE and HP). A simple comparison at the same temperature is not very relevant because daily consumption depends on the average temperatures of the previous days. The use of 2 models poses a problem because the HP model has a much lower explanatory power than the DE model due to the much lower heating consumption, which has less influence on daily consumption for all uses and potentially larger confident interval. Finally, direct modelling of energy savings along the whole metered period is also possible by including the effect of HP use as a qualitative variable in a quanti-quali type model.

### Tipping points

Daily consumption as a function of outside temperature (Figure 4) enables us to determine several key values: daily consumption gradient, base consumption (non-thermosensitive) and non-heating temperature. To estimate the non-heating temperature ( $T_{threshold}$ ), we used an optimized linear regression of the daily consumption  $C_d$  explained by the average outside temperature of the day and the previous 2 days:

$$C_d = \beta_0 + \beta_1 * T_d + \beta_2 * T_{d-1} + \beta_3 * T_{d-2} + \varepsilon \text{ with } T_d < T_{threshold} \quad \text{eq.1}$$

The  $T_{threshold}$  is chosen to provide the best goodness of fit (i.e., maximum  $R^2$ ) using a Monte-Carlo procedure ( $T_{threshold}$ ) = [14:22]) (Leschinski 2019) of a linear modelling using lm function (R Core Team 2023). To calculate this temperature, we only use the outside temperature, as this is the most important variable (Figure 7). This non-heating temperature ( $T_{threshold}$ ) is used to filter the data so that only consumption dependent on the outside temperature is modelled in this study.

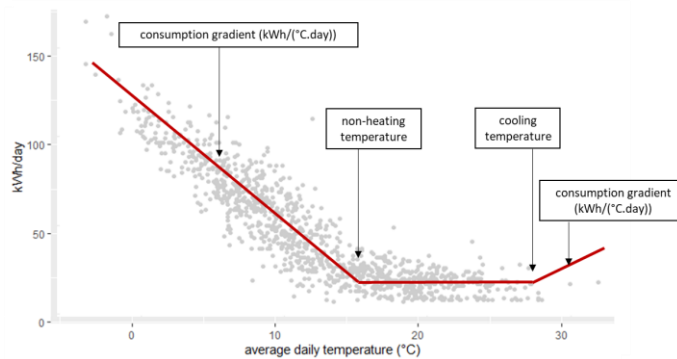


Figure 4. Schematic representation of the daily consumption gradients (all end-uses), non-heating temperature ( $T_{threshold}$ ) and cooling temperature (tipping points).

### Modelling the reference situation

If we refer to the IPMVP's (International Performance Measurement and Verification Protocol) methods of measurement and verification (M&V), to calculate energy savings, we are in the framework of option C (whole facility) based on smart meter data (EVO 2018). Energy savings are assessed based on before and after situation, with an adjusted baseline assessing what would have happened if the energy efficiency measure had not been introduced (reference situation).

Several methods have been used to estimate energy savings by modelling the before situation based on a general framework of regression modelling based on various predictors as we supposed (Genuer et al. 2008):

$$Y = s(X_i) + \varepsilon \quad \text{eq.2}$$

with  $Y$ : variable of interest (numerical response),  $X_i$ : explanatory variables (predictors),  
 $s$ : regression function,  $\varepsilon$ : error term.

Currently, a range of more or less complex methods are readily available and usable beyond the simplest linear regression model. In this study, we test a set of methods to determine whether one method is particularly more efficient than the others. Certain methods used in this study are now commonly used to forecast the electrical load curve, readers are invited to refer to (Hong and Fan 2016, Wang et al. 2019) for a full overview of the methods used. We should note that a time frame of more than 1 year correspond to long term forecasting (Hong and Fan 2016). According to (Hong and Fan 2016), models can be classified according to different categories:

- non- interpretable model (black box) vs. interpretable models (white box),
- univariate models (no explanatory variables) vs. multivariate models (using weather information).

Complex methods (including use of different algorithms) are also used to forecast consumption and to identify the most important drivers (Jiao et al. 2022). Artificial Neural Network are also used (Sarmas et al. 2022). It should be borne in mind that forecasting methods based on sensor data within the building, like in (Edwards et al. 2012), are outside the scope of this study.

The 11 methods used in this study are:

- Linear regression (ordinary least square - OLS) (R Core Team 2023) which minimizes the sum of the squared error by finding the coefficients of a linear equation.
- Quantile Regression (QR) of median value (i.e.  $\tau = 0.5$ ) (Koenker 2022). The quantile regression allows the coefficients to differ from one quantile to another. Compared to modelling the mean (OLS), QR is more appropriate in the presence of extreme values (outliers) or when the distribution of the variable of interest is very spread out (Givord and D'Haultfœuille 2013).
- Partial Least Square Regression (PLSR) (Liland et al. 2023). PLS regression was designed to accommodate collinearity between linearly dependent explanatory variables (Mevik and Wehrens 2023), this is particularly the case for the temperature variable from one day to another or other weather variables.
- Enhanced Adaptive Regression Through Hinges (EARTH) (Milborrow 2023). Non-parametric regression model based on the Multivariate Adaptive Regression Splines in (Friedman 1991) combining recursive partitioning and spline fitting.
- Support Vector Machine (SVM) (Meyer et al. 2023). Regression with kernel-based learning methods using hyperplane that minimizes the deviation of data points from the predicted values (Smola and Scholkopf 2004).
- Random Forest (RF) (Liaw and Wiener 2002). The principle of random forests is to combine independent decision trees built using several bootstrap samples and choosing randomly at each node a subset of explanatory variables. The final prediction is the average of these individual tree predictions with no consideration for the errors of other trees (Genuer et al. 2008).
- Extreme Gradient Boosting (XGBoost) (Chen et al. 2023). A sequence of decision trees, with each new tree correcting the errors made by the previous ones. At the contrary to RF each subsequent tree focusing on correcting the mistakes of the ensemble up to that point (Chen and Guestrin 2016).
- Generalized Additive Models (GAM) (Hastie 2023). Models using unspecified multiple smooth additive functions (Hastie and Tibshirani 1990).
- Quantile GAM (qGAM) (Fasiolo et al. 2021). Fitting additive quantile regression models based on splines (Fasiolo et al. 2020).
- Prophet (Taylor and Letham 2021). A time series forecasting model with explanatory variables and 3 mains component: trend, seasonality weekly, yearly and holidays like a generalized additive model (Taylor and Letham 2018). It should be noted that only this model models consumption by taking time into account (date as explanatory variable).
- Staking model (Superlearner) (Poley et al. 2021). A generalized stacking ensemble learning technique estimating the performance of multiple models creating an optimal weighted average of those models and finds the combination of algorithms minimizing the cross-validated risk (Poley and van der Laan 2010).

To produce the models, we attempt to explain daily consumption ( $C_d^{DE}$ ) using DE space heating, as a function of the following explanatory variables:

- $T_d, T_{d-1}, T_{d-2}$ : average external temperature for days  $d, d-1$  and  $d-2$  to consider inertia.
- $T_{min}, T_{max}$ : to consider extrema in temperature during the day  $d$ .
- $WS$  and  $WD$ : effect of the wind (speed and direction) linked to the permeability of the dwelling.

- *RH*: relative humidity.
- *P*: atmospheric pressure.

We must notice that T, P and RH are slightly correlated and  $T_d$ ,  $T_{min}$  and  $T_{max}$  are highly correlated (Figure 5).

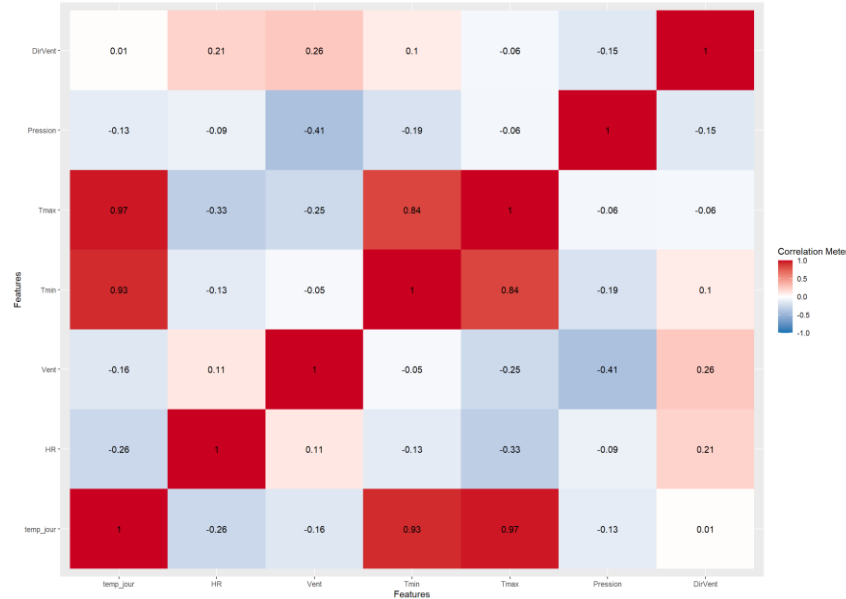


Figure 5. Correlation analysis of explanatory variables.

To train and to evaluate models, we use a training sample (80% of the data by random selection) and the complementary test sample (20%). The variables between the two samples are presented in Table 2 and show similar values. It should be noted that the link between temperature and time is not completely broken by the random sampling because we integrate the outdoor temperature over 2 days backwards as variables.

To estimate the performance of the models on the test sample, we use the Mean Absolute Percentage Error (MAPE), the MAE (Mean Absolute Error) and the Root Mean Squared Error (RMSE) supplemented by the R-Squared (Coefficient of Determination) regression Score ( $R^2$ ) as calculated by (Yan 2022).

Table 2. Train and test samples of the before situation (with  $T_d < T_{threshold}$ )

Sample	Train				Test				Unit	Code
	min	median	mean	max	min	median	mean	max		
Daily consumption	7.00	55.00	56.45	168.00	8.00	55.00	55.39	165.00	Day	C
External temperature	-3.85	9.304	9.117	16.96	-3.408	9.467	9.229	16.946	kWh	T
Relative humidity	49.0	87.0	85.4	100.0	49.0	86.0	85.5	98.0	%	RH
Wind Speed	2.717	12.915	14.277	41.80	3.408	14.050	15.996	44.317	km/h	WS
Wind Direction	23.91	195.21	186.45	346.67	27.92	205.42	194.41	343.75	degree	WD
Atmospheric pressure	976.5	1018	1016.9	1045.2	985.2	1016.9	1017.2	1040.7	hPa	P

### Energy savings assessment

Energy savings ( $ES$ ) are calculated by comparing observed consumption ( $C_{observed}^{HP}$ ) (all end uses) from the smart meter with that modelled from the previous situation (i.e. the modelled consumption ( $C_{modelled}^{DE}$ ), if the DE heating were still in use (reference situation)), based on actual data for the observed period (2021-06 to 2023-08):

$$ES = \sum_d \widehat{C_{modelled}^{DE}} - \sum_d C_{observed}^{HP} = \sum_d (\widehat{C_d^{DE}} - C_d^{HP}) = \sum_d ([s(X_t^d) + \varepsilon] - C_d^{HP}) \text{ eq.3}$$

In this study, we simply calculate the energy savings between a modelled situation and a real situation, but it would be possible for future work to directly produce a model of energy savings as a function of the explanatory variables analysed.

We have assessed the uncertainty associated with the  $ES$  estimate by calculating confidence intervals associated with this estimate. The 95% confidence interval is represented by the 2.5% and 97.5% level empirical quantiles of a set of  $M = 1,000,000$  energy savings simulations. Each simulation  $j = 1 \dots M$  is calculated as the sum of the daily savings simulations (i.e. 700 days from 2021-06 to 2023-08 with  $T_d < T_{threshold}$ ), assuming independence

from one day to the next. Each daily savings simulation is carried out by a random sampling based on the inversion of the empirical distribution function provided by the EARTH method.

## Results

### *Metered consumption and daily consumption gradient*

Figure 6 shows the daily consumption as a function of the outdoor temperature and clearly demonstrates the difference in slope for the thermal gradient between the before and after situations. The non-heating temperature  $T_{threshold}$  with DE space heating is estimated at 16.96°C ( $R^2=0.84$ ). The use of data filtering according to the non-heating temperature estimated by a simple linear regression model (eq.1) may be questionable as we are using more complex models capable of handling strong inflections. The decision to keep the filtering is based on the idea of comparing models, some of which (e.g. OLS, QR) are not capable of handling these inflections.

The baseline consumption (i.e. with  $T_d > T_{threshold}$ ) in the before situation is on average 17 kWh/day with a standard deviation of 7 kWh (min = 7 kWh, max=49 kWh).

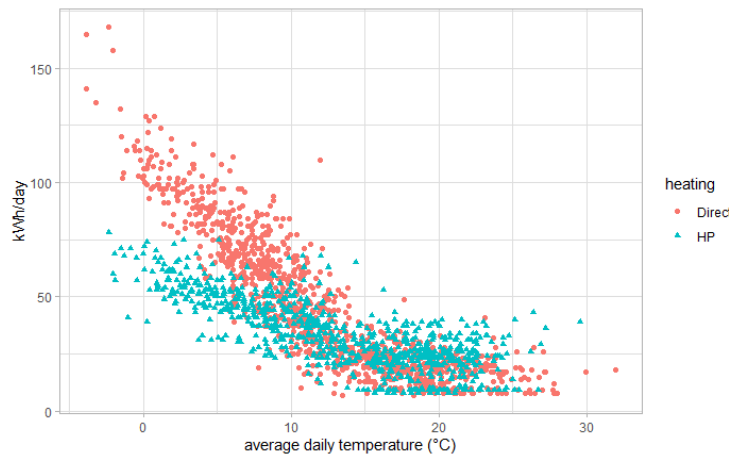


Figure 6. Daily consumption (kWh, all end-uses) depending on the average outdoor temperature ( $T_d$  in °C) and the type of space heating equipment (circle: DE, triangle: HP) over the metered period (2018-06 to 2023-08).

Average daily consumption changes after installation of the HP as well as the standard deviation is lower after the installation of the HP compared to the previous situation with direct electric heating (Table 3).

Table 3. Daily electric consumption (all end-uses). Q1, Q3: 1st and 3rd quartile, sd: standard deviation

situation	min	Q1	median	mean	Q3	max	sd
DE	7	20	38	46	68	168	30.2
HP	8	23	32	34	44	78	15.2

### *Models' performance*

The performance of the models calculated is presented in order of performance in Table 4 using all the explanatory variables. As a result, the models are not optimised by the choice of significant explanatory variables (which could be endogenous for some model) and can affect their performance. But this is a methodological approach to automated data processing. All models are run using the default parameters. It can be noted that the RMSE of the OLS model based solely on temperature over 3 days and that with all the variables are respectively 12.11 kWh and 11.52 kWh ( $R^2=0.824$  vs.  $R^2=0.847$ ). The choice of a simple model is open to discussion.

We should also note that the ranking of the models depends on the choice of the non-heating temperature. Here, as we have an approach based on linear regression, this does not give any particular advantage to models capable of handling inflections.

According to MAPE, MAE and RMSE, with the exception of Prophet, whose results are slightly better in terms of RMSE and  $R^2$ , the best performing model is the EARTH model followed by qGAM and SVM. However, the difference in performance between these models remains small for all models (<10% compared to the lowest MAPE), except for OLS and QR (>20%). It should be noted that some performing models are not explanatory models (SuperLearner, RF). The simple and explanatory models (OLS, QR) performed the worst.



We should note that, as noted by (Hong and Fan 2016), model performance is highly data dependent, and it is difficult to conclude that one model is generally superior to another. The importance of this last point in assessing energy savings will be discussed later.

In addition, depending on the choice of the non-heating temperature, the performance of certain models changes because some are better able to deal with discontinuities. Our method of determining the temperature limits discontinuities by construction.

**Table 4. performance of models (on sample test) ranked by increasing MAPE and increasing RMSE.**

Model	MAPE	RMSE	MAE	R <sup>2</sup>	Type of model
Earth	0.197	10.057	7.880	0.879	interpretable
SVM	0.199	10.867	8.883	0.858	non-interpretable
qGAM	0.200	10.058	7.884	0.879	interpretable
RF	0.204	10.456	8.043	0.869	non-interpretable
GAM	0.205	10.235	8.064	0.874	interpretable
Superlearner	0.209	10.325	8.883	0.872	non-interpretable
XGBoost	0.210	10.927	8.606	0.857	non-interpretable
Prophet	0.211	10.009	8.045	0.880	interpretable
PLSR	0.233	11.286	8.560	0.847	interpretable
OLS	0.240	11.308	8.749	0.847	interpretable
QR	0.244	11.519	8.823	0.841	interpretable

It should be noted that some models are better at handling extreme values, such as GAM (Table 5). Some models give minimum values that are too low (PLSR, OLS, QR, deviation >50%). For maximum values, the differences are much smaller (<20%) for all models. This shows the complexity of considering the transition from temperature-sensitive consumption to non-sensitive consumption and very low-consumption due to absence from home (Figure 3).

**Table 5. Analysis of the consumption  $C_d^{DE}$  (in kWh/day) predicted by the models based on the test sample.**

Model	min	1 <sup>st</sup> Quartile	median	mean	3 <sup>rd</sup> Quartile	max
<i>Test sample</i>	8.00	28.50	55.00	55.90	79.50	165.00
Earth	12.80	31.33	54.27	55.33	75.41	137.10
SVM	13.50	32.23	53.12	55.73	75.98	147.12
qGAM	11.43	31.88	53.39	55.86	76.52	145.08
RF	15.75	33.57	53.66	55.70	75.33	146.94
GAM	12.32	37.76	53.69	55.90	76.14	151.24
Superlearner	10.53	33.67	53.53	55.91	76.19	148.19
XGBoost	12.75	30.07	53.26	53.57	72.68	132.81
Prophet	6.63	37.01	55.83	55.64	74.82	133.18
PLSR	-2.33	37.84	55.20	55.84	75.77	133.18
OLS	2.09	29.14	57.20	54.18	74.75	129.14
QR	1.81	37.309	55.80	56.24	75.99	136.22

In the following sections, we detailed the best performing model by presenting interesting outputs and issues encountered. The EARTH model is based on using hinge function (Milborrow 2023):

$$\hat{f}(x) = \sum_{i=1}^k a_i [\pm(X_i - c)]_+ + b + \epsilon_i \text{ eq.4}$$

with: c: knot, b: intercept, a<sub>i</sub>: coefficients

The EARTH model considers 8 variables out of 9 ( $T_{min}$  is considered by the model as a non-significant variable like in many models used in this study). The importance of the variables is as follows:  $T_d$ ,  $T_{max}$ ,  $T_{d-2}$ ,  $WS$ ,  $T_{d-1}$ ,  $P$ ,  $WD$ ,  $RH$ . The effect of the significant variables of the outcome of EARTH model is presented in Figure 7. One of the important interests of the earth model (appendix) is that it is interpretable and relies on simple linear regressions. Nevertheless, a physical interpretation of the effect of the variables on consumption remains to be done in certain cases.

Note that if we consider only outdoor temperatures for 3 days as explanatory variables, the EARTH model performs similarly than the worst model (RMSE=11.669, MAE= 8.718, MAPE=0.209 and  $R^2 = 0.837$ ).

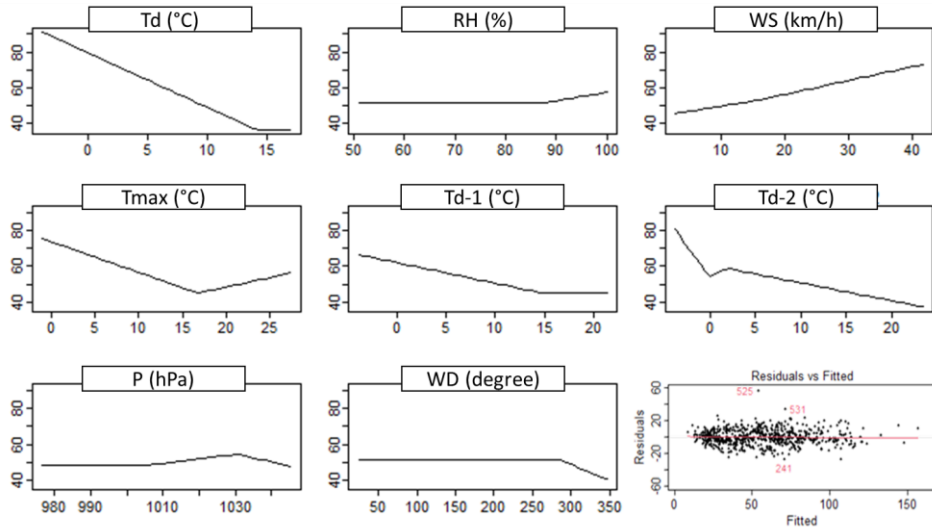


Figure 7. plots for each variable included in EARTH model. The x-axis contains the explanatory variable values, and the y-axis contains the effect on the consumption values (in kWh/day) (calculations according to (Milborrow 2022)). Bottom-right residual vs. fitted values according to (Milborrow 2023).

### Assessment of energy savings and discussion

Observed consumption (from 2021-06 to 2023-08) for the period following the installation of the HP was 20,597 kWh including all end-uses and the PHEV consumption (Figure 8). Using the values of the explanatory variables, the models calculated a reference consumption (as it was in the before situation) and estimated the gross energy savings by comparison based on equation 3 (Table 7). The net energy savings should be higher if the consumption of the PHEV is deducted over the period. Moreover, the ES should be reduced if the potential air-conditioning consumption in summer is added (specific to our case study which installs a reversible HP) by taking into account the period with  $T_d > T_{threshold}$ .

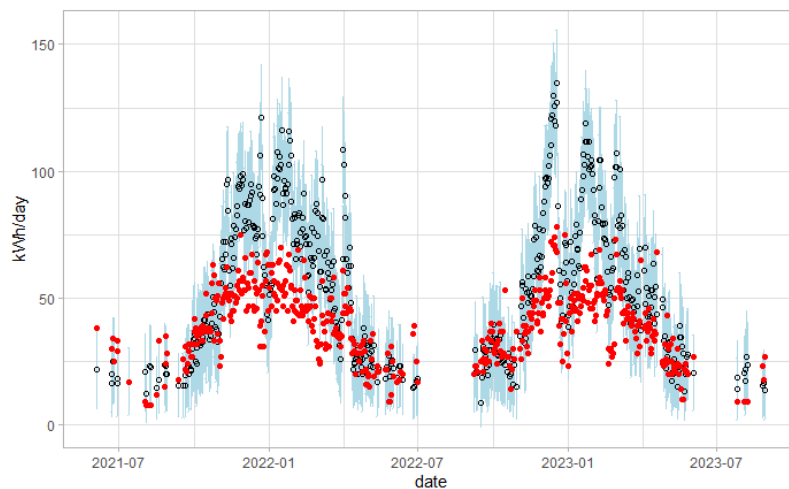


Figure 8. daily consumption (kWh/day, all end-uses) following HP implementation (black hollow: predicted DE consumption according to EARTH model, full red: observed consumption with HP, pale blue: prediction interval [5%-95%]). Absence of data for days with  $T_d > T_{threshold}$ .

Despite significant models' performance differences: MAPE ranging from 2% to 36% compared to the best model, the impact on the estimated gross energy savings remains low (<5%) if we also take the best model (EARTH) as the reference. In addition, there does not appear to be any link between MAPE or RMSE values and estimation values.

**Table 7. Estimated energy savings (in kWh from 2021-06 to 2023-08, see Figure 9) based on the different models.**

Model	Gross energy savings (kWh)	Deviation from the best model
EARTH	8,109 ( $\pm 480$ )	Reference (Ref. $\pm 6\%$ )
SVM	8,036	-0.90%
qGAM	7,840	-3.32%
RF	8,162	0.66%
GAM	7,985	-1.52%
Superlearner	8,077	-0.39%
XGBoost	7,996	-1.39%
Prophet	8,237	1.58%
PLSR	7,831	-3.43%
OLS	8,079	-0.37%
QR	8,026	-1.02%

We note that the estimated gross energy savings of the different models (from -3.3% to 1.6% compared to EARTH result) (Table 7) remains within the uncertainty of the EARTH model calculation ( $\pm 6\%$ ) (Figure 9) indicating an insignificant difference between the different results.

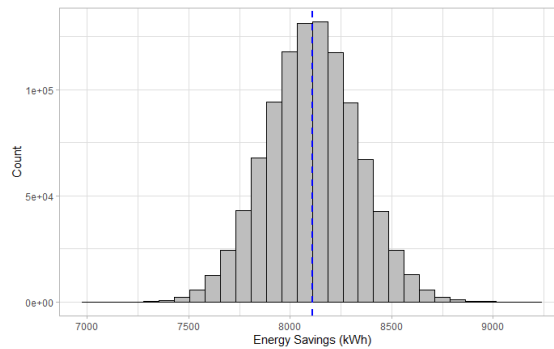


Figure 9. Distribution of the gross energy savings estimated with the model EARTH using uncertainty calculation ( $M=1,000,000$ ) (mean = 8,109 kWh, [5%: 7,533 kWh - 95%: 8,478 kWh]).

In conclusion, even with relatively simple models, and not necessarily the best performing ones, it is possible to estimate gross energy savings from daily consumption (due to large error compensation). In addition, the quality of some models may be questionable as they do not always meet the necessary criteria (heteroscedasticity, variance inflation factor, etc.). The choice of a model with no strong theoretical constraints is therefore an interesting one.

However, estimating net energy savings remains complex without information on changes in other consumption (PHEV in our case study and air conditioning) or behaviour. To correctly estimate the net energy savings, we need to look at the outdoor temperature sensitivity of the PHEV's consumption, due in particular to pre-conditioning, the use of the vehicle's heating system and variations in battery capacity. This will be the subject of future developments.

## Conclusion

In this study, we used a case study to test different methods of estimating energy savings based on daily consumption of all end-uses from a smart meter before and after renovation (i.e. installation of a heat pump). While we are aware that no general conclusions can be drawn from this case study, it does provide a basis for considering the feasibility of deploying it on a wider scale to assess the potential of energy savings.

The algorithms used vary in performance but, as we are aggregating daily consumption over a long period (at least 1 year), these differences in performance are not reflected in the estimated gross energy savings due to error compensation. The major problem does not therefore lie in the choice of the algorithm.

The approach developed in this study applies notably to both electricity and gas consumption (i.e. daily consumption from a smart meter). In addition, it should be noted that in the case of gas consumption, the outside temperature sensitivity of the overall consumption is greater because the smart meter only measures 3 end-uses at best (space heating, DHW, cooking) and does not include appliances on the contrary of electricity. The higher the proportion of measured consumption accounted for by space heating, the more efficient the models. This suggests that small-scale retrofit on heating consumption (e.g. installation of double-glazed windows) would be more difficult to estimate.

The use of data consumption from smart meters raises questions about how to process it to estimate the net energy savings from gross savings. Numerous factors/biases are likely to modify the net savings compared to the gross savings (before/after situation): rebound effect, modification of the initial situation after renovation both in terms of the building (e.g. extension of space area like converted attic, new end-use like in our case study PHEV) and the household (e.g. number of people, time spent using the home). One approach to addressing this question would be to analyse the evolutions of basic (non-thermosensitive) consumption.

The case-by-case approach is therefore difficult to apply in a residential context and a good estimate of energy savings from smart meter data remains complex. A panel approach seems preferable to reduce the uncertainties associated with unknown changes. Thus, a large panel should make it possible to obtain energy savings evaluation that is as close as possible to the actual energy savings calculated using a statistically representative panel based on data from smart meters and field surveys.

## References

- Binet G., Girardot D., Denoual M., Bontemps S. (2021). *A techno-explicit model implementation for long term electrical load curve prospective*, CIRED 2021, paper number 994
- CEREN (2023). Centre d'Études et de Recherches Économiques sur l'Énergie, <https://www.ceren.fr/>
- Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T, Li M, Xie J, Lin M, Geng Y, Li Y, Yuan J (2023). *xgboost: Extreme Gradient Boosting*. R package version 1.7.6.1
- Chen T., Guestrin C. (2016). *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD international conference on knowledge Discovery and Data Mining. August 2016, p785-794
- Chen, T., He T., Benesty M., Khotilovich V., Tang Y., Cho H., Chen K., Mitchell R., Cano I., Zhou T., Li M., Xie J., Lin M., Geng Y., Li Y., Yuan J. (2023). *xgboost: Extreme Gradient Boosting*. R package version 1.7.6.1
- CSTB (2022). *Fiche typologie - Pavillon à fermettes et combles aménagés (1974 à aujourd'hui)*. <https://programmeproofeel.fr>
- Edwards R., New J., Parker L. (2012). *Predicting Future Hourly Residential Electrical Consumption: A Machine Learning Case Study*. Energy and Buildings
- EHPA (2022). European Heat Pump Association. *REPowerEU – Doubling heat pump sales requires a “Heat Pump Accelerator”*. Press Release, March 2022
- Enedis, (2022). <https://www.enedis.fr/le-compteur-linky-un-outil-pour-la-transition-ecologique>
- Fasiolo M, Wood SN., Zaffran M., Nedellec R., Goude Y. (2021). *qgam: Bayesian Nonparametric Quantile Regression Modeling in R*. Journal of Statistical Software, 100(9), 1-31
- Fasiolo M., Wood S. N., Zaffran M., Nedellec R., Goude Y. (2020). *Fast calibrated additive quantile regression*. Journal of the American Statistical Association, 116, 1402–1412
- Friedman (1991). *Multivariate Adaptive Regression Splines (with discussion)* Annals of Statistics 19/1, 1–141 <http://projecteuclid.org/euclid.aos/1176347963>
- Genuer R., Poggi J-M., Tuleau C. (2008). *Random Forests: some methodological insights*. INRIA, rapport de recherche. ISSN 0249-6399, 35p
- Givord P., D'Haultfœuille X. (2013). *La régression quantile en pratique*. INSEE, No M2013/01
- Hastie T (2023). *gam: Generalized Additive Models*. R package version 1.22-2
- Hastie T., Tibshirani R. (1990). *Generalized additive models*. Chapman & Hall/CRC
- Historique Météo (2023). <https://www.historique-meteo.net>
- Hong T., Fan S. (2016). *Probabilistic electric load forecasting: A tutorial review*. International Journal of Forecasting, 32, 914–938
- IEA (2022). *The Future of Heat Pumps*. International Energy Agency, World Energy Outlook Special Report
- EVO (2018). Generally Accepted M&V Principles. International Performance Measurement and Verification Protocol, Efficiency Valuation Organization, 20p
- Jiao j., Behrisch M., Brugger H., Eichhammer W. (2022). *Identifying drivers of residential energy consumption by explainable energy demand forecasting*. ECEEE Summer Study Proceedings 533-539
- Koenker R (2022). *Quantreg: Quantile Regression*. R package version 5.94

Leschinski C. (2019). *Monte Carlo: Automatic Parallelized Monte Carlo Simulations*. R package version 1.0.6

Liaw A., Wiener M. (2002). *Classification and Regression by randomForest*. R News 2(3), 18-22

Liland K, Mevik B, Wehrens R (2023). *\_pls: Partial Least Squares and Principal Component Regression\_*. R package version 2.8-3

Meteostat (2022). <https://meteostat.net/fr/>

Mevik B-H., Wehrens R. (2023). *The pls package: Principal component and partial least squares regression in R*. Journal of Statistical Software, 18(2):1–24, 2007

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F (2023). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*, TU Wien. R package version 1.7-13

Milborrow S. (2022). *plotmo: Plot a Model's Residuals, Response, and Partial Dependence Plots*. R package version 3.6.2

Milborrow S. (2023). *earth: Multivariate Adaptive Regression Splines*. R package version 5.3.2

Osso D., Laurent M-H. (2017). *Radiographie des logements « tout élec » construits avant les années 2000*. CVC - Chauffage, Ventilation, Conditionnement d'air, n°897, 38-41

Poley E., van der Laan M. (2010). *Super Learner in prediction*. University of California, Berkeley U.C. Berkeley Division of Biostatistics. Working Paper Series. Paper 266  
<https://biostats.bepress.com/cgi/viewcontent.cgi?article=1269&context=ucbbiostat>

Polley E, Le Dell E, Kennedy C, van der Laan M (2021). *SuperLearner: Super Learner Prediction*. R package version 2.0-28

R Core Team (2023). *R: A language and environment for statistical computing*. Version 4.3.2. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>

RTE (2024). *Europe : une sécurisation de la trajectoire de neutralité carbone reposant sur une forte électrification de la demande et une accélération du déploiement des renouvelables*. Chap. 5, 61p., <https://assets.rte-france.com/prod/public/2024-01/2024-01-02-bilan-previsionnel-2023-chapitre-5-europe.pdf>

Sarmas E., Martinakis V., Doukas H., Dimitropoulos N., Zucika A. (2022). *Monitoring the impact of energy conservation measures with Artificial Neural Networks*. ECEEE 2022 proceedings, 571-579

Smola A., Scholkopf B. (2004). *A tutorial on support vector regression*. Statistics and Computing 14: 199–222, 2004

Taylor S., Letham B. (2018). *Forecasting at Scale*, The American Statistician, 72:1, 37-45

Taylor S., Letham B. (2021). *Prophet: Automatic Forecasting Procedure*. R package version 1.0

Thomaßen G., Kavvadia K., Jimenez Navarro J-P. (2021) *The decarbonisation of the EU heating sector through electrification: A parametric analysis*. Energy Policy, 148, 111929

Uniclimate (2024). Bilan 2023 et perspectives 2024 des industries thermiques, aéraliques et frigorifiques

Wang Y., Chen Q., Hong T., Kang C. (2019). *Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges*. IEEE Transactions on Smart Grid.

Yan Y. (2016). *MLmetrics: Machine Learning Evaluation Metrics*. R package version 1.1.1

## Appendix

**EARTH model of the daily consumption (all end-uses,  $T_d < T_{\text{threshold}}$ ) in the before situation (DE space heating). Selected 14 of 19 terms (additive model), and 8 of 9 predictors.**

Hinge (variable)	coefficients	Hinge (variable)	coefficients
(Intercept)	11.2492178	$h(T_{d-2} - 0.0458333)$	9.1524809
$h(14.1417 - T_d)$	3.0373880	$h(1.95 - T_{d-2})$	6.8835855
$h(RH - 87)$	0.4560831	$h(T_{d-2} - 1.95)$	-10.1428098
$h(15.3958 - WS)$	-0.5633174	$h(\text{Pression} - 1005.59)$	0.2474842
$h(WS - 15.3958)$	0.7640152	$h(\text{Pression} - 1030.54)$	0.6989871
$h(16.8 - T_{\text{max}})$	1.7121654	$h(WD - 286.25)$	-0.1805410
$h(T_{\text{max}} - 16.8)$	1.0648267	$h(\text{Pression} - 1005.59)$	0.2474842
$h(14.5792 - T_{d-1})$	1.1503995		