Data-driven algorithms to model space heating in households equipped with gas boilers and smart thermostats

Gerard Mor[a,b,*], Jordi Cipriano[a,b], Mariano Tur[c], Eloi Gabaldón[b], Benedetto Grillone[d], Daniel Chemisana[e]

[a]*Department of Environmental and Soil Sciences & University of Lleida, Rovira Roure 191, 25198 Lleida, Spain*
[b]*Centre Internacional de Mètodes Numèrics a l'Enginyeria. Building Energy and Environment Group. CIMNE-Lleida. Pere de Cabrera 16. Office 2G. 25001. Lleida. Spain*
[c]*BAXI.BDR-Thermea. Salvador Espriu, 9, 08908, L'Hospitalet de Llobregat, Barcelona, Spain.*
[d]*Centre Internacional de Mètodes Numèrics a l'Enginyeria. Building Energy and Environment. GAIA building (TR14). Rambla Sant Nebridi 22. 08222 Terrassa. Spain*
[e]*Department of Environmental and Soil Sciences & INSPIRES, University of Lleida, Rovira Roure 191, 25198 Lleida, Spain.*

Abstract

In this paper

Keywords: Space heating, energy efficiency, Modelling, Genetic Algorithm

Nomenclature

LPF Low-Pass Filter function

$RMSE$ Root Mean Squared Error

SH Space Heating

1. INTRODUCTION

In 2017, households represented 27 % of final energy consumption, or 17 % of gross inland energy consumption, in the EU [1]. The main use of energy by households was for space heating, 64 % of final energy consumption [1]. Most EU Member States rely mainly on natural gas and electricity for meeting theses needs, followed by renewable energies, mostly solid bio fuels. This high dependence in natural gas clearly determines any strategy to achieve the binding targets of increasing at least 32.5 % energy efficiency by 2030 [2]. To achieve this increase in the energy efficiency, several low carbon technologies are already available in the market: electric heat pumps, hybrid heat pumps combined with fossil-fuel boilers, or district heating networks are successful examples on how to pave the way to decarbonize the space heating sector. Many research studies have been focused in demonstrating their cost effectiveness and how these technologies can increase the energy efficiency in several European countries [3, 4, 5, 6, 7]. However, although the technologies are readily available, the involvement and participation of end users is completely necessary. They must be part of the solution, and this can only be achieved if the drivers of this low-carbon transition can find new ways of interacting with end users to facilitate their empowerment and their active participation in the European energy markets.

*Corresponding author
Email addresses: gmor@cimne.upc.edu (Gerard Mor), cipriano@cimne.upc.edu (Jordi Cipriano), mtur@cimne.upc.edu (Mariano Tur), egabaldon@cimne.upc.edu (Eloi Gabaldón), bgrillone@cimne.upc.edu (Benedetto Grillone), daniel.chemisana@udl.cat (Daniel Chemisana)

The unfolding of these new user driven energy services needs some kind of advanced metering infrastructure (AMI) or a massive adoption of smart home devices. To date, Member States have committed to rolling out close to 200 million smart meters for electricity and 45 million for gas by 2020 at a total potential investment of 45 billion [8]. By 2020, it is expected that almost 72 % of European consumers will have a smart meter for electricity while 40 % will have one for gas.

On the other hand, for the last years we have seen a fast penetration of the emerging Internet of Things (IoT) technologies into residential homes. Nowadays, smart devices are inevitable in our lives [9, 10]. Smart thermostats are one of them.This smart thermostats allow to remotely control the home climate, to show the temperature and energy consumption in real time or to communicate with intelligent cloud-base IT systems to incorporate self-learning capabilities. These are crucial features to accommodate efficient techniques to increase the efficiency of HVAC systems and decrease energy costs. However, studies have shown that more than half of the installed thermostats are used in manual modes, due to the inefficiency of automatic settings [11, 12]. Moreover, it was found that 57 % of households were occupied nearly all the time,limiting the potential energy savings in non-occupied hours.

Therefore, while the thermostats capabilities to control temperature are well understood, less is known about their effectiveness to enable savings. The uncertainty in these savings is increasingly important because manufacturers are adding many new features and functions that affect the ability and ease of saving energy. To reduce this uncertainty in evaluating the potential savings of smart meters or smart thermostats, advanced and dynamic energy performance models, able to consider not only the weather dependent variables (outdoor temperature and humidity), but also the internal temperature and, what it is even more important, the set point temperature, are fully necessary. These dynamic energy models must be derived from the available measured variables, which are mainly temperatures (indoor, outdoor and set point) and energy consumption. They are needed not only at aggregated level but also at individual level. Individual modeling provides a lot more information than modeling the total consumption of group of users, including information of direct practical interest, e.g. about inter-individual variability, as well as information which is useful for interactive and user based energy awareness programs.

In [13], a review of the state of the art of models to predict natural gas consumption, from 2000 to 2010, was presented. An exponential increase of papers was detected, specially in the lower forecasting area level (regional, gas distribution and individual). The predominant trend of these research works was a combination of optimizing tools with more classic forecasting models. After 2010, several authors continued using statistical and stochastic methodologies to predict and characterize aggregated gas consumption of residential or commercial groups of buildings [14, 15]. At individual level, in [16] a non linear mixed effect model (NLME) is developed for the prediction of the individual gas consumption at daily basis. After comparing the results with time series models (TS), such as ARX and ARMAX, they concluded they perform similarly but having their merits and problems. The NLME are cleaner and clearer, while ARX and ARMAX are better for local adaptation to sudden and abrupt changes within a single individual. In [17],linear (ARX) and nonlinear (neural networks and SVM) are applied to forecast natural gas consumption in a daily basis. The solar radiation as an exogenous variable was included in the models and the accuracy improved.This research work performed a very detailed evaluation of several TS models in non-occupied test homes and clearly quantified the model accuracy improvement by introducing the solar radiation as an exogenous variable, however these test conditions are very far from real and occupied buildings where the heating system is thermostatically controlled by the user through the set point temperature. In [18], a step wise calibration of a dynamic thermal empirical model of a residential building, was performed. The calibration included some user dependant parameters, such as the air ventilation rates, however, the constraints derived by the set point temperature control were not included in the analysis. More recently, [19] developed a dynamic home thermal model built upon the standard RC (Resistor-Capacitor) approach and tested it with data from a test home in free floating situation. This RC model includes the effect of most of the exogenous variables, such as the internal and external temperatures, the wind and the solar radiation, however it doesn't consider the effect of the set point temperature and the user behaviour.

Several research works have studied the impact of installing smart thermostats. However, the majority of them were centered in the problem of control and were focused on occupancy modelling. In [20], occupancy sensors were installed in 8 homes to automatically turn off the HVAC system in sleeping or non-occuppied hours. []

In this paper, a new methodology to evaluate the energy consumption of households equipped with thermostatic controlled space conditioning systems is developed. The methodology is based on data driven models derived from the information gathered by smart thermostats. It includes all the involved variables such as the outdoor temperature, the solar radiation, the wind velocity and direction, the indoor temperature, the space heating power consumption and the set point temperature. The methodology is validated in real cases within the winter season, however,the results and conclusions are applicable also to space cooling system as long as they are controlled by thermostatic control variables. The methodology develops statistical learning models able to capture the household energy performance in all the modes driven by the control variable, the set point temperature. The mode when the indoor temperature is higher than the set point threshold is modelled by a first regression model where the indoor temperature is the dependent variable and the space heating power consumption is one of the input variables. This space heating power consumption becomes a dependent variable, fed by the indoor temperature and other exogenous variables, when the indoor temperature is lower than the set point temperature threshold. Both regression models are combined to forecast the expected energy consumption when a certain set point temperature schedule is applied.

The paper starts with a mathematical description of the regression models and of the inputs variables transformation. It follows with a description of the processes to train both models and to optimise the regression parameters. The procedure to combine the two regression models to predict the energy consumption due to a certain set point temperature schedule is then described. The paper finishes with the application of the methodology over a set of households of north eastern Spain, which are equipped with condensed gas boilers driven by smart thermostats.

2. Methodology

The energy performance of a household is influenced by the indoor and outdoor conditions, by the physical and geometric characteristics of the building, by the type of space conditioning system, and by the thermostatic control, which in most cases is a thermostat managed by the occupants. Therefore, when modelling the energy performance, all these variables should be considered and included in the models. In this paper, a methodology which combines two regression models, named the demand-side and the supply-side models, is developed to accurately predict the energy consumption of thermostatic-controlled heating systems when the occupants change the set point temperature schedules. The statistical learning models used to describe the dynamical performance of both the household and the heating system are Auto-Regressive linear models with exogenous variables (ARX). These impulse response models can describe these time-varying processes in a fast and efficient way, however a previous transformation of the input variables is recommended to improve the parameters fitting and the model accuracy.

2.1. Demand-side model

The demand side model is defined by an ARX model represented by the indoor temperature (T^i) as the dependent variable. A set of exogenous variables and the space heating power consumption are the input variables. This model captures how the heat flows out of the building and how the indoor temperature is affected by the heating system. The model formula is described in equation 1.

$$\begin{aligned} \phi(B)y_t &= \hat{y}_t + \varepsilon_t \\ \phi(B)T^i_t &= \omega_h(B)\Phi^h_t + \omega_e(B)T^{e,lp}_t + \omega_w(W^s_t \times W^d_t) + \omega_s(I^{sol}_t \times S^{az}_t) + \varepsilon_t \end{aligned} \tag{1}$$

The coefficients ω_h, ω_e, ω_w and ω_s are the parameters of the model. The auto-regressive term $\phi(B)$ is defined in equation 2, where: n is the number of lags decided after the parameters optimization and B is the backward shift operator $B^k_{y_t} = y_{t-k}$.

$$\phi(B) = 1 + \phi_1 B^1 + ... + \phi_n B^n \quad (2)$$

The input variables considered in the model of equation 1 are:

1. Time lagged indoor temperatures (T^i) to characterize the inertia of the building.
2. Low-pass filtered outdoor temperature ($T^{e,lp}$) to characterize the energy loses through the envelope of the building due to changes in the outdoor temperature. Additionally, the fourth power of this temperature ($T^{e,lp4}$) is used to represent the radiative transfer of the building envelope.
3. Energy consumption of the heating system (Φ^h) to characterize the increase of the indoor temperature.
4. Solar irradiance (I^{sol}) and sun azimuth (S^{az}) to characterize the solar gains to the building.
5. Wind speed (W^s), wind direction (W^d) and difference between indoor and outdoor temperature (Ψ) to characterize heat losses due to air infiltration and convection through the envelope.

2.2. Supply-side model

This model estimates the amount of energy needed to warm the household considering the base load of the space heating system, the indoor temperature $T^{i,lp}$, the outdoor temperature $T^{e,lp}$ and the solar position P^{sol}. This model only accounts for the amount of energy for space heating when there is a need for heating, so the model is fitted only with $\Phi^h_t > 0$.

$$\Phi^h_t = \alpha_s + \omega_i \Theta(T^{i,lp}_t, T^{i,lp}_{t-1}) + \omega_e \Psi(\begin{bmatrix} T^{i,lp}_t \\ T^{i,lp}_{t-1} \end{bmatrix}, \begin{bmatrix} T^{e,lp}_t \\ T^{e,lp}_{t-1} \end{bmatrix}) + \omega_{sol} P^{sol}_t + \varepsilon_t \quad (3)$$

Besides applying low-pass filtering over the indoor and outdoor temperatures some additional transformed terms are considered. $\Theta(x, y)$ (equation 4) is the difference between the indoor temperature at current time step and at previous time step, which represents the thermal inertia of the household. Its related model parameter ω_i quantifies the amount of energy needed to gain temperature inside the household. The term $\Psi(x, y)$, expressed in equation 5, is the difference between the averaged indoor temperature and the averaged outdoor at current and previous time steps. Therefore, the related model parameter ω_e represents an equivalent thermal heat transfer coefficient of the household.

$$\Theta(x, y) = \begin{cases} x - y & if\ (x - y) > 0 \\ 0 & otherwise \end{cases} \quad (4)$$

$$\Psi(x, y) = \overline{x} - \overline{y} \quad (5)$$

In the case of the solar position P^{sol}, each element is defined by a vector formed by the solar azimuth angle A^{sol} and the solar height H^{sol} at each time step, as expressed in equation 6. The model parameter ω_s represents an equivalent solar gains coefficient. The independent, term α_s, quantifies the base or permanent space heating load of the household.

$$P^{sol}_t = \begin{bmatrix} A^{sol}_t \\ H^{sol}_t \end{bmatrix} \quad (6)$$

2.3. Transformation of input variables

2.3.1. Low-pass filter of exogenous variables

The Low-Pass Filter (lp)allows to transform the exogenous variables used as input to the models into variables that better represent the dynamics of the system to enhance the fit of the model. It assumes that the dynamics of the buildings can be described by lumped parameter RC models, see for example [21, 22]. This assumption means the response in the consumption of the heating system, at least the proportion due to the heating or cooling system, to changes in the climate exogenous variables, can be modelled as a first order low pass filter. Based on this assumption, it is reasonable to apply low-pass filters to all the exogenous variables in order to eliminate the high input frequencies that might negatively affect the model training. The discrete time implementation of this first order RC low-pass filter is the exponentially weighted moving average of each variable with the filter parameter tuned to match the response of the building to each effect separately:

$$x^{lp} = LPF(x, \alpha) \tag{7}$$

$$x_t^{lp} = \alpha x_t + (1 - \alpha) x_{t-1}^{lp} \tag{8}$$

Where: x^{lp} is the filtered exogenous variable; α is the filter parameter $[0, 1]$; and x is the original time series of the exogenous variable.

In this paper, the indoor temperature $T^{i,lp} = LPF(T^i, \alpha_i)$ and the outdoor temperature $T^{e,lp} = LPF(T^e, \alpha_e)$ are the exogenous variables which are low-pass filtered before being included as input variables of the two models.

2.3.2. Infiltrations ($W^s\Phi$)

2.4. Models training and parameters optimization

The parameters of the input variables α_i and α_e are estimated using a genetic optimization procedure defined in algorithm 1

. The fitness function considered in this optimization is defined in equation 9. This function is only evaluated for the time steps included within the validation period, which means the input variables are different than the ones used in the training period. The evaluation procedure of each chromosome consists in the following steps: The model fitting of equations 1 and 3 is performed using the iterative re-weighted least squares method (IWLS) [25]

- Decode the chromosome in gray code to a float representation of the α_i and α_e parameters.
- Calculate the $T^{i,lp}$ and $T^{e,lp}$ input variables of the models.
- Train the space heating power consumption as output model (equation 3) using the training period data.
- Predict the validation period data.
- Calculate the of this prediction.

$$RMSE_v = \sqrt{\frac{1}{n_v} \sum_{vt=1}^{n_v} (\widehat{\Phi_{vt}^h} - \Phi_{vt}^h)^2} \tag{9}$$

Once the optimized α_i and α_e are estimated, they are used in both models, as this parameters corresponds to the physical characterization of the building.

2.5. Simulation of space heating control scenarios

Both regression models are used in a combined way to to simulate the energy consumption due to a certain set point temperature schedule. This models combination provides a very flexible procedure to evaluate energy saving scenarios based on reducing the set point temperature in winter or increasing it in summer period. It is also a very useful method to evaluate the amount of energy available for a flexible interaction with the gas or electricity markets. Algorithm 2 describes the procedure to simulate the effect of modifying the control variable, the set point temperature, schedule over both the indoor temperature and the space heating power consumption.

Algorithm 1: Genetic Algorithm

Data: Historical weather and consumption data for one household
Result: Optimal α_i and α_e of the household
DEFINE a training and validation periods over the initial data set;
CHOOSE an encode method to use a binary genetic algorithm;
INITIALISE population with random candidate solutions;
ENCODE each candidate solution to a gray code representation;
EVALUATE each candidate solution; $i = 0$;
while $i \leq MaxIteration$ **do**
  SELECT individuals for the next generation;
  RECOMBINE pairs of parents;
  MUTATE the resulting offspring;
  ENCODE each candidate solution to a gray code representation;
  EVALUATE each candidate solution; $i = i + 1$;
DECODE the individual with minimum $RMSE_v$;
return optimal α_i, α_e;

3. CASE STUDY

3.1. Case study data sets

A real practice of the whole methodology is experienced over a test pilot case which is formed by 100 households placed in the western north area of Spain. Each household is equipped with a condensed gas boiler which is controlled by a smart thermostat. Both the condensing boiler and the smart thermostat, named BAXI Connect, are manufactured and provided by the company BAXI. Figure 1 shows a set of pictures of the process of installing this smart thermostat and of setting it up to control the space heating and domestic hot water distribution systems of one of the involved households. The smart thermostat follows the Open Therm communication protocol [23] to communicate towards the gas boiler and Wireless connection to communicate towards the household router. The variables communicated by the thermostat are: indoor temperature, set point temperature, outdoor temperature (boilers equipped with an extra sensor), approximate space heating heat power consumption, and approximate domestic hot water heat power consumption. These data are communicated every 15 minutes and the measurement tolerance corresponds to 2 KWh for the space heating power consumption and 0.5 C for the temperature readings. The testing period started in December 2018 and finished in May 2019. However, due to data protection and privacy issues,the involved customers were sequentially activated. Only from March to May a representative number of customers was achieved, therefore, the analysis performed in this research is limited to this time period. Because of data protection and privacy issues, no information about the household address and of its main building features is available. Only the postal code is accessible by the gathering system. The IT architecture of this case study is formed by the local smart thermostat which transfers all the data to a central server managed by BAXI. These data is pseudo annonymized and it is communicated, through a REST Full API communication layer, to the big data analytics cloud. The details of the these distributed and big data processing framework were described in detail in [24].

3.1.1. Climate data

Some households,which have an extra temperature sensor placed at the outern skin of the household, provide data of the climate dependent exogenous variables (outdoor temperature and relative humidity), however, for all the other cases,the outdoor temperature, the wind speed and the wind bearing data are obtained from a weather web service of the company Dark Sky [25]. These climate data are based on the approximate location of each household. Additionally, the global incident solar radiation in a planar surface is obtained from the Copernicus European Union's Earth observation program [26], which entails to model more accurately the solar heat gains of the household. [27] suggest that the incident solar radiation can be indirectly considered by including the solar angles as an input for the models.

Algorithm 2: Algorithm for the simulation of space heating power consumption ($\widehat{\Phi_t^h}$) over a defined time period $t \in [1, j]$ considering a virtual setpoint temperatures scenario ($T_t^{s,sim}$).

The fitted model of the indoor temperature (Needed to estimate $\widehat{T_t^{i,lp}}$); the fitted model of the space heating power load (Needed to estimate $\widehat{\Phi_t^h}$); the autoregressive order for the indoor temperature model (n); the hysteresis of the thermostat (H^{therm}); the last indoor temperatures ($T_{-k}^{i,lp} \forall k \in 1, ..., n$) and the space heating power load of the boiler (Φ_{-1}^h) before the evaluation period (both are needed to fix the initial conditions of the household); the real space heating power load (Φ_t^h) and weather features ($T_t^{e,lp}, S_t^{pos}$) over the evaluation period; and the setpoint temperature scenario ($T_t^{s,sim}$) to be simulated during the evaluation period $t \in [1, j]$ The estimated energy savings $\Phi_{savings}^h$ over actual consumption by setting a setpoint temperature scenario ($T_t^{s,sim}$) defined in the algorithm input. **begin**

```
DEFINE the A input-output matrix (A ∈ ℝ^{j,8+n}). The column names are:
  T_t^{e,lp}, T̂_t^{i,lp}, T̂_{t-k}^{i,lp} ∀k ∈ [1,n], Θ_t*, Ψ_t*, Φ̂_t^h, Φ̂_{t-1}^h, A_t^{sol}, H_t^{sol};
  *Θ_t, Ψ_t corresponds to Θ(T̂_t^{i,lp}, T̂_{t-1}^{i,lp}) and Ψ([T̂_t^{i,lp}; T̂_{t-1}^{i,lp}], [T_t^{e,lp}; T_{t-1}^{e,lp}]), respectively;
SET A_{1,Φ̂_{t-1}^h} = Φ_{-1}^h ;                      // Initial autoregressive conditions
for k = [1,n] do
    for l = [0,n-k] do
        SET A_{k+l,T̂_{t-(k+l)}^{i,lp}} = T_{-k}^{i,lp};
SET t = 1 ;                                           // Iterator of timesteps
while t ≤ j do
    SET A_{t,Φ̂_t^h} = 0;
    ESTIMATE A_{t,T̂_t^{i,lp}} using the indoor temperature model and A_{t,*};
    if A_{t,T̂_t^{i,lp}} < (T_t^{s,sim} - H^{therm}) then     // Heat is needed
        SET A_{t,T̂_t^{i,lp}} = T_t^{s,sim} + H^{therm};
        CALCULATE A_{t,Θ_t}, A_{t,Ψ_t} using, among others, last set A_{t,T̂_t^{i,lp}};
        ESTIMATE A_{t,Φ̂_t^h} using the space heat power consumption model and A_{t,*};
        ESTIMATE A_{t,T̂_t^{i,lp}} using the indoor temperature model and A_{t,*};
    A_{t+1,Φ̂_{t-1}^h} = A_{t,Φ̂_t^h} ;                  // Update autoregressive conditions
    for k = [1,N] do
        SET A_{t+k,T̂_{t-k}^{i,lp}} = T̂_t^{i,lp};
    SET t = t + 1 ;                                   // Next timestep
Φ_{savings}^h = ((Σ_{t=1}^n Φ_t^h - Σ_{t=1}^n A_{t,Φ̂_t^h}) / Σ_{t=1}^n Φ_t^h) × 100%
```

return the energy savings $\Phi_{savings}^h$; the setpoint temperatures scenario $T_t^{s,sim}$; and the A matrix, which contains all the intermediate results of the simulation;

4. RESULTS

In Fig. 2 the variables used the demand side and supply models of one of the involved households, are shown. The historical period used to train the models includes 1^{st} March to 31^{st} May 2019. Starting from the top, the dark-green line corresponds to the set point temperature, the dark orange line corresponds to the indoor temperature, the violet line corresponds to the outdoor temperature, the magenta line corresponds to the space heating power consumption, the light green line corresponds to the direct normal incident solar radiation, the dark yellow

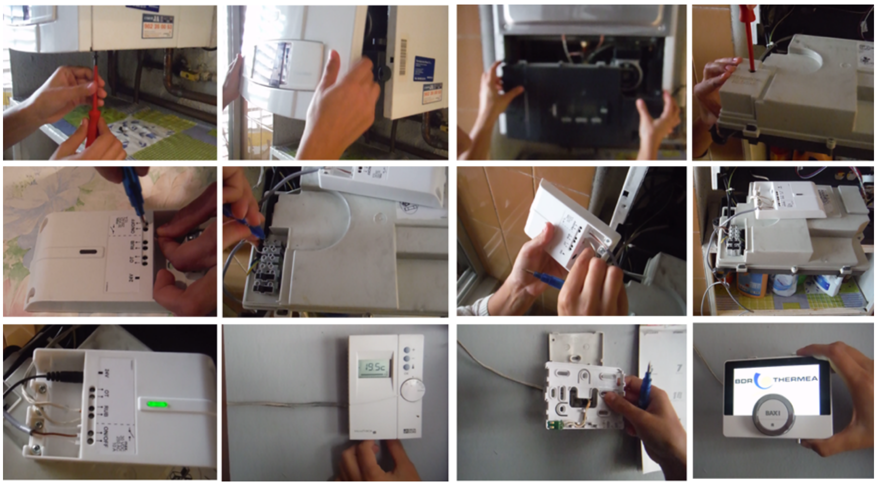

Figure 1: Pictures of the installation of the smart thermostat (BAXI CONNECT) in of the case study household

line correspond to the wind speed times the difference between indoor and outdoor temperature, the brown line corresponds to the difference between the set point temperature and the indoor temperature and the dark grey line corresponds to the difference between the outdoor and the indoor temperatures. The validation period is not shown in Fig. 2 because it is formed by a random selection of 10 days between February to May. These validation days were excluded from the training period.

The Fig. 3 shows the fitted regression parameters of the supply model (space heating power consumption as output) of one involved household .

The Fig. 4 shows the residual analysis of the model of one household and the heat power as output, for the training period. As it can be observed, the white noise is achieved for the residuals distribution, which means the model is considered as validated. The Fig. 5 shows the fitted coefficients of the model of one household in the case of considering the internal temperature as the output The Fig. 6 shows the residual analysis of the model of one household and the indoor temperature as output, for the training period. As it can be observed, the white noise is achieved for the residuals distribution, which means the model is considered as validated.

Once the two models are fitted, they could be used to estimate the energy consumption when different set point temperature schedules are considered. Fig. 7 depicts the real energy consumption, the indoor temperature and set point temperature of a household during March 2019, and a simulated set point temperature scenario using the characterization models of the household. As can be seen, the $T^{s,sim}$ pattern is similar to the real one T^{s} but assigning lower temperatures to the reduced set point temperature, from 19 to 18°C, and the comfort temperature, from 22 to 20°C. The estimated energy savings of the change in set point temperature schedule are a 16.58% less of energy consumption than the real one (528 kWh) during March 2019.

5. CONCLUSIONS

The presented methodology is based on statistical techniques to infer physical features of the building

General considerations and limitations of the approach: The presented methodology is highly dependent on the availability of the input data and the complexity and quality of the desired output.

Future work: the results can be combined with location-related information, such as weather, administrative social or economic data to assess the energy performance of buildings

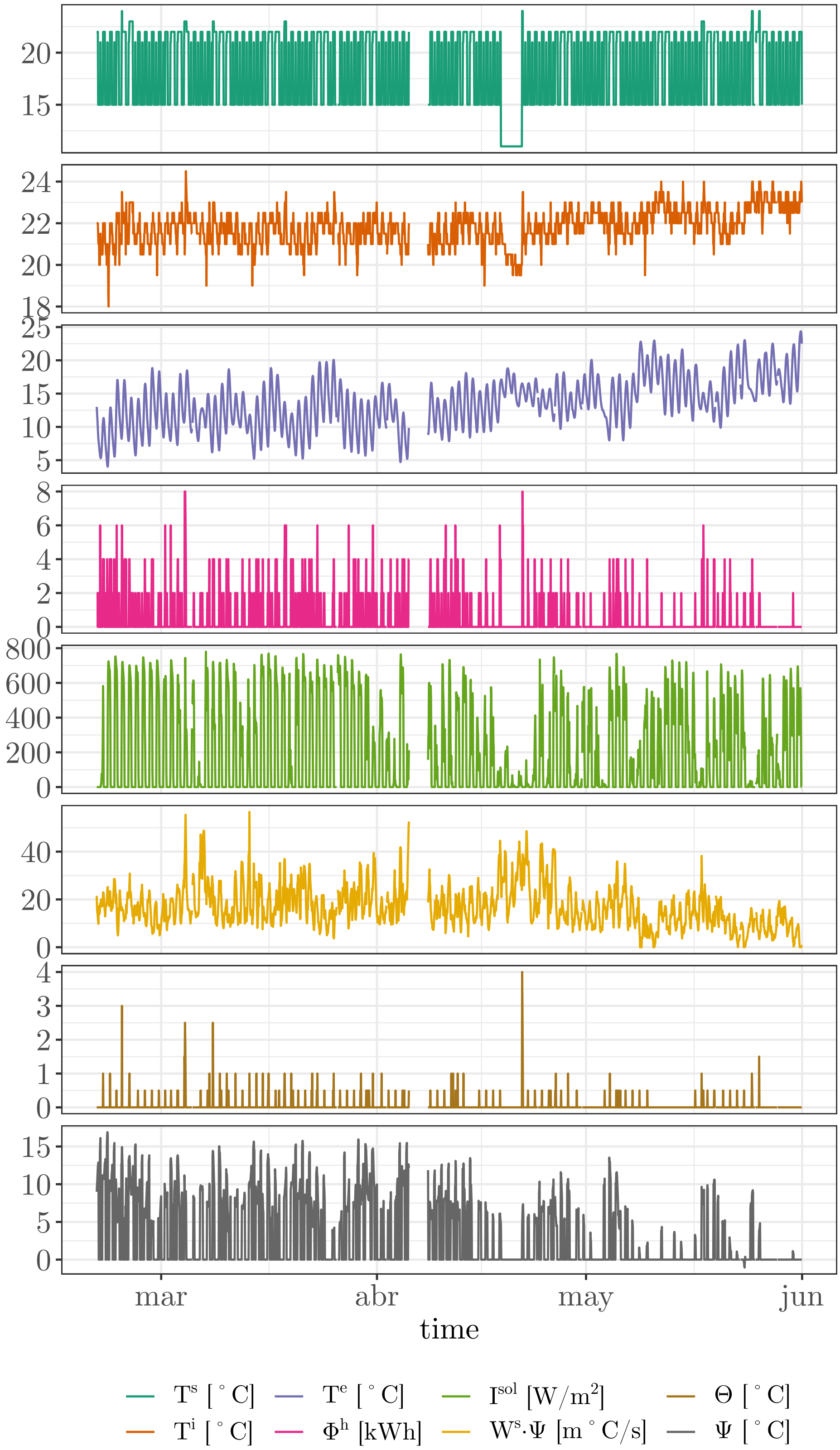

Figure 2: Input and output variables of the demand side and supply models of one household. Period:1^{st} March to 31^{st} May.

6. ACKNOWLEDGMENT

This work was developed during the project COMRDI15-1-0036 within the RIS3CAT community of the Catalan government. We thank all the BAXI employees who provided insight

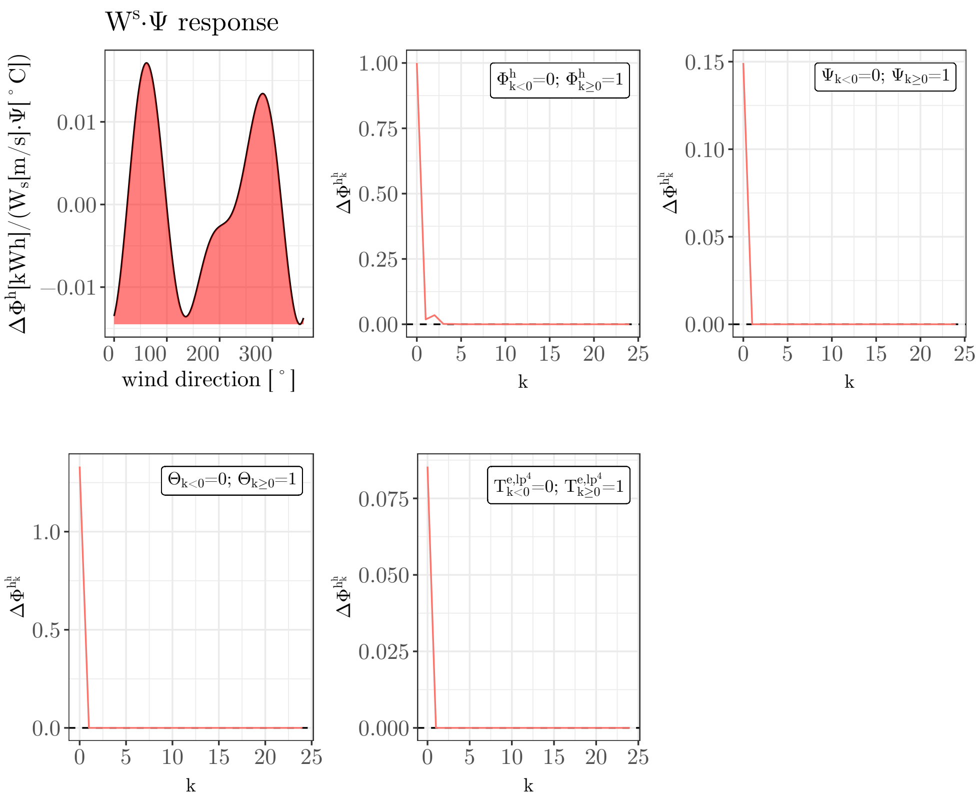

Figure 3: Coefficients of the supply model (Φ^h as output)

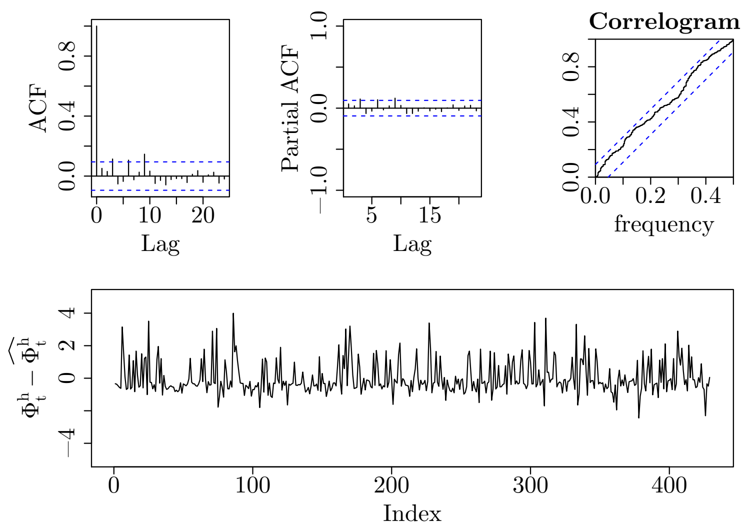

Figure 4: Training residuals of the model with Φ^h as output

and expertise that greatly assisted the work, and all their clients that participated in the project.

References

[1] Eurostat. *Energy consumption in households - Statistics Explained*. May 2019. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_consumption_in_households.

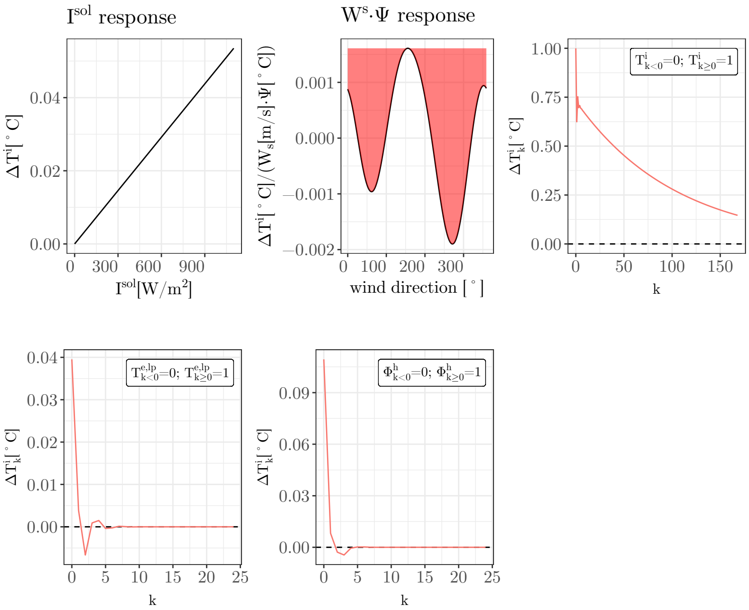

Figure 5: Coefficients of the model with T^i as output

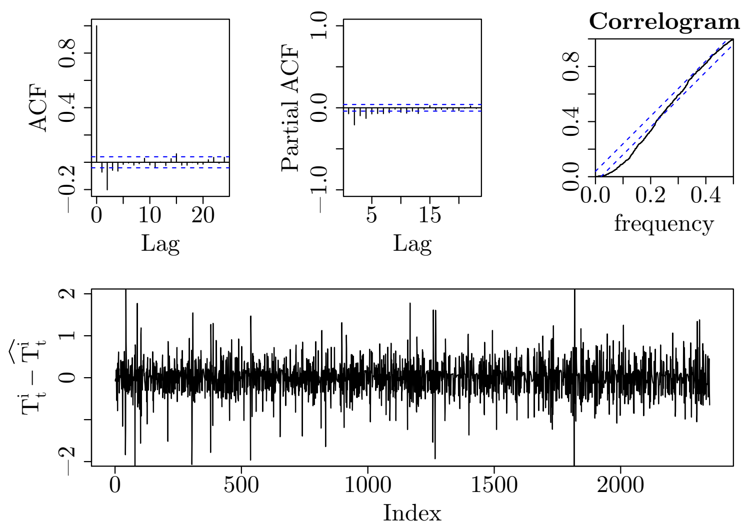

Figure 6: Training residuals of the model with T^i as output

[2] Directorate-General for Energy (European Commission). *Clean energy for all Europeans*. 2019. URL: https://publications.europa.eu/en/publication-detail/-/publication/b4e46873-7528-11e9-9f05-01aa75ed71a1/language-en?WT.mc_id=Searchresult&WT.ria_c=null&WT.ria_f=3608&WT.ria_ev=search.

[3] Anne Stafford. "An exploration of load-shifting potential in real in-situ heat-pump/gas-boiler hybrid systems". In: *Building Services Engineering Research and Technology* 38.4 (2017), pp. 450–460. DOI: 10.1177/0143624416688727. URL: https://doi.org/10.1177/0143624416688727.

[4] Matteo Rivoire et al. "Assessment of Energetic, Economic and Environmental Performance of Ground-Coupled Heat Pumps". en. In: *Energies* 11.8 (Aug. 2018), p. 1941. DOI: 10.3390/en11081941. URL: https://www.mdpi.com/1996-1073/11/8/1941 (visited on 08/06/2019).

[5] Xi Zhang et al. "Economic assessment of alternative heat decarbonisation strategies through coordinated operation with electricity system fffdfffdfffd UK case study". In:

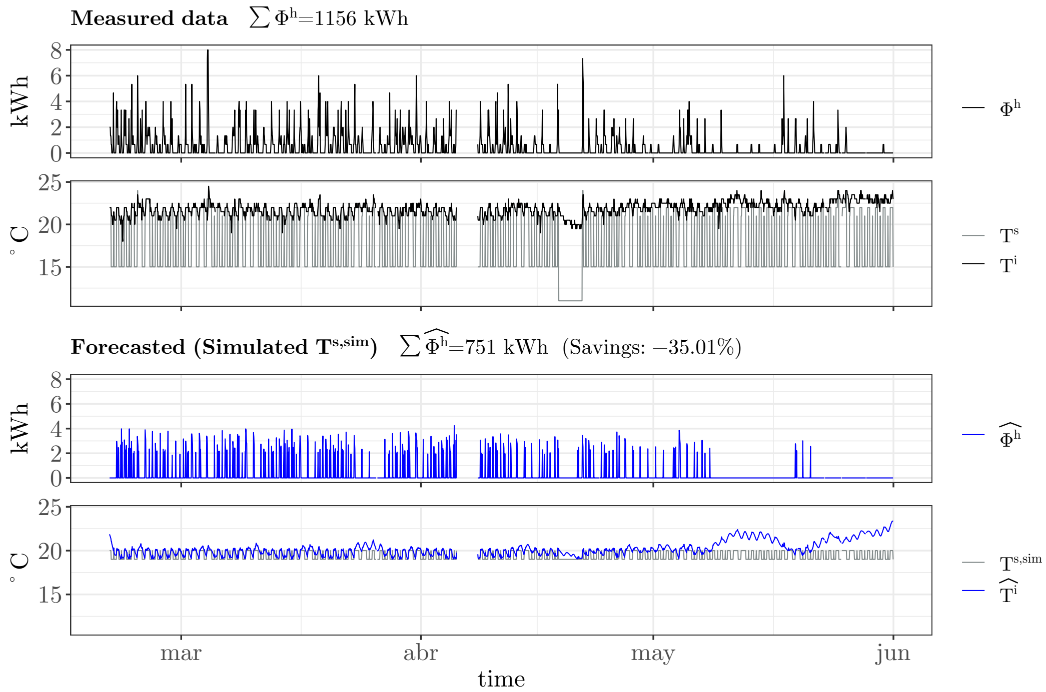

Figure 7: Simulation of $T^{s,sim}$

Applied Energy 222 (July 2018), pp. 79–91. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2018.03.140. URL: https://www.sciencedirect.com/science/article/pii/S0306261918304781 (visited on 08/06/2019).

[6] Stephen Clegg and Pierluigi Mancarella. "Integrated electricity-heat-gas modelling and assessment, with applications to the Great Britain system. Part II: Transmission network analysis and low carbon technology and resilience case studies". In: *Energy* (Feb. 2018). ISSN: 0360-5442. DOI: 10.1016/j.energy.2018.02.078. URL: http://www.sciencedirect.com/science/article/pii/S0360544218303062 (visited on 08/06/2019).

[7] M. Jarre et al. "Opportunities for heat pumps adoption in existing buildings: real-data analysis and numerical simulation". In: *Energy Procedia*. Sustainability in Energy and Buildings 2017: Proceedings of the Ninth KES International Conference, Chania, Greece, 5-7 July 2017 134 (Oct. 2017), pp. 499–507. ISSN: 1876-6102. DOI: 10.1016/j.egypro.2017.09.608. URL: http://www.sciencedirect.com/science/article/pii/S1876610217347422 (visited on 08/06/2019).

[8] European Commission-COM(2014) 356 final. *REPORT FROM THE COMMISSION-Benchmarking smart metering deployment in the EU-27 with a focus on electricity*. 2014. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014DC0356&from=EN.

[9] H. Farhangi. "The path of the smart grid". In: *IEEE Power and Energy Magazine* 8.1 (Jan. 2010), pp. 18–28. ISSN: 1540-7977. DOI: 10.1109/MPE.2009.934876.

[10] Gary Newe. "Delivering the Internet of Things". In: *Network Security* 2015.3 (Mar. 2015), pp. 18–20. ISSN: 1353-4858. DOI: 10.1016/S1353-4858(15)30020-9. URL: http://www.sciencedirect.com/science/article/pii/S1353485815300209 (visited on 08/08/2019).

[11] Marco Pritoni et al. "Energy efficiency and the misuse of programmable thermostats: The effectiveness of crowdsourcing for understanding household behavior". en. In: *Energy Research & Social Science* 8 (July 2015), pp. 190–197. ISSN: 22146296. DOI: 10.1016/j.erss.2015.06.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S2214629615000730 (visited on 08/08/2019).

[12] Therese Peffer et al. “How people use thermostats in homes: A review”. In: *Building and Environment* 46.12 (Dec. 2011), pp. 2529–2541. ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2011.06.002. URL: http://www.sciencedirect.com/science/article/pii/S0360132311001739 (visited on 08/08/2019).

[13] Bofffdfffdidar Soldo. “Forecasting natural gas consumption”. In: *Applied Energy* 92 (Apr. 2012), pp. 26–37. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2011.11.003. URL: http://www.sciencedirect.com/science/article/pii/S0306261911006982 (visited on 08/05/2019).

[14] Ehsan Tavakoli and Nader Montazerin. “Stochastic analysis of natural gas consumption in residential and commercial buildings”. In: *Energy and Buildings* 43.9 (Sept. 2011), pp. 2289–2297. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2011.05.017. URL: http://www.sciencedirect.com/science/article/pii/S0378778811002313 (visited on 08/05/2019).

[15] Longquan Diao et al. “Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation”. In: *Energy and Buildings* 147 (July 2017), pp. 47–66. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2017.04.072. URL: http://www.sciencedirect.com/science/article/pii/S037877881731469X (visited on 08/05/2019).

[16] Marek Brabec et al. “A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers”. In: *International Journal of Forecasting*. Energy Forecasting 24.4 (Oct. 2008), pp. 659–678. ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2008.08.005. URL: http://www.sciencedirect.com/science/article/pii/S0169207008000976 (visited on 08/05/2019).

[17] Bofffdfffdidar Soldo et al. “Improving the residential natural gas consumption forecasting models by using solar radiation”. In: *Energy and Buildings* 69 (Feb. 2014), pp. 498–506. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2013.11.032. URL: http://www.sciencedirect.com/science/article/pii/S0378778813007299 (visited on 08/05/2019).

[18] Wancheng Li et al. “Stepwise calibration for residential building thermal performance model using hourly heat consumption data”. In: *Energy and Buildings* 181 (Dec. 2018), pp. 10–25. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2018.10.001. URL: http://www.sciencedirect.com/science/article/pii/S0378778818324757 (visited on 08/05/2019).

[19] Junke Wang et al. “Predicting home thermal dynamics using a reduced-order model and automated real-time parameter estimation”. In: *Energy and Buildings* 198 (Sept. 2019), pp. 305–317. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2019.06.002. URL: http://www.sciencedirect.com/science/article/pii/S0378778818338350 (visited on 08/05/2019).

[20] Jiakang Lu et al. “The Smart Thermostat: Using Occupancy Sensors to Save Energy in Homes”. In: *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. SenSys ’10. event-place: Zürich, Switzerland. New York, NY, USA: ACM, 2010, pp. 211–224. ISBN: 978-1-4503-0344-6. DOI: 10.1145/1869983.1870005. URL: http://doi.acm.org/10.1145/1869983.1870005.

[21] Henrik Aalborg Nielsen and Henrik Madsen. “Modelling the heat consumption in district heating systems using a grey-box approach”. en. In: *Energy and Buildings* 38.1 (Jan. 2006), pp. 63–71. ISSN: 03787788. DOI: 10.1016/j.enbuild.2005.05.002. URL: http://linkinghub.elsevier.com/retrieve/pii/S0378778805000745 (visited on 11/27/2018).

[22] Peder Bacher et al. “Short-term heat load forecasting for single family houses”. en. In: *Energy and Buildings* 65 (Oct. 2013), pp. 101–112. ISSN: 03787788. DOI: 10.1016/j.enbuild.2013.04.022. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378778813002752 (visited on 01/14/2020).

[23] *OpenTherm Protocol Specification-version 2-2*. Feb. 2003. URL: https://www.domoticaforum.eu/uploaded/Ard%20M/Opentherm%20Protocol%20v2-2.pdf.

[24] G. Mor et al. "EMPOWERING, a Smart Big Data Framework for Sustainable Electricity Suppliers". In: *IEEE Access* 6 (2018), pp. 71132–71142. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2881413.

[25] Powered by Dark Sky. *Dark Sky API*. URL: https://darksky.net/poweredby/.

[26] Council of the European Union , European Parliament. *Regulation (EU) No 377/2014 of the European Parliament and of the Council of 3 April 2014 establishing the Copernicus Programme and repealing Regulation (EU) No 911/2010 Text with EEA relevance*. Apr. 2014. URL: https://publications.europa.eu/en/publication-detail/-/publication/976616e8-cb7c-11e3-b74e-01aa75ed71a1.

[27] Dan Kelley et al. *Solar Angle Function. OCE R Package*. June 2019. URL: https://cran.r-project.org/web/packages/oce/oce.pdf.