Grzegorz DUDEK, Marcin JANICKI

Czestochowa University of Technology, Department of Electrical Engineering

doi:10.15199/48.2017.03.02

Nearest Neighbour Model with Weather Inputs for Pattern-based Electricity Demand Forecasting

***Abstract**. A nearest neighbour model with exogenous variables representing weather factors for electricity demand forecasting in short horizons is proposed. Weather factors are included into the k-nearest neighbours regression model as forecast pattern contexts. Similarities between contexts affect the weights assigned to the patterns in the regression model. The proposed model is examined in several forecasting problems with different levels of influence of weather factors on the demand. For strong influence the forecast results are improved due to incorporation of weather inputs.*

***Streszczenie.** Zaproponowano model najbliższych sąsiadów ze zmiennymi egzogenicznymi reprezentującymi czynniki pogodowe do krótkoterminowego prognozowania zapotrzebowania mocy. Czynniki pogodowe wprowadzone są do modelu regresji k-najbliższych sąsiadów jako konteksty obrazów prognoz. Podobieństwa pomiędzy kontekstami wpływają na wagi obrazów w modelu regresyjnym. W badaniach symulacyjnych obserwuje się poprawę rezultatów dzięki wprowadzeniu kontekstów, gdy wpływ czynników pogodowych na zapotrzebowanie jest istotny. (**Model najbliższych sąsiadów z wejściami pogodowymi do opartego na obrazach prognozowania zapotrzebowania na moc**).*

Keywords: short-term load forecasting, nearest neighbour model, pattern similarity-based forecasting.
Słowa kluczowe: prognozowanie krótkoterminowe obciążeń, model najbliższego sąsiedztwa, prognozowanie oparte na podobieństwie obrazów.

Introduction

Nearest neighbour methods such as *k*-nearest neighbours (*k*-NN) are ones of the most popular methods of pattern recognition. Despite their simplicity they successfully compete with more sophisticated methods of classification and regression and are even known to be the most effective ones [1]. Reduction of the recognition region to the neighbourhood of the input sample (this neighbourhood is represented by a set of *k* nearest neighbours of the input sample) introduces local character of the model. It is worth noting that *k*-NN methods do not require any assumptions about domain and data representation except that there is specified some measure of similarity or distance [2]. Also any assumptions about codomain of the target mapping are not needed. If codomain is a set of real numbers, the algorithm solves the problem of function approximation. When codomain is a set of categories, the *k*-NN solves data classification problem.

In our case the problem of electricity demand forecasting is treated as a regression problem. The regression function estimator has the nonparametric form:

$$\widehat{m}(\mathbf{x}) = \sum_{i=1}^{N} w(\mathbf{x}, \mathbf{x}_i)\mathbf{y}_i \tag{1}$$

where: $\mathbf{x}$ is the vector of predictors, $\mathbf{y}$ is the response vector, N is the number of learning samples and $w(.)$ is the weighting function.

A vector $\mathbf{x}$ consists of the lagged values of the forecasted variable, while a vector $\mathbf{y}$ consists of its future values. So the model maps previous values of some variable into its future values. The weighting function $w(.)$ assigns positive weights to learning samples which belong to the set of *k*-nearest neighbours of the input vector $\mathbf{x}$. Samples outside of this set have zero weights. According to (1) the forecasted y-vector is calculated as the weighted mean of the learning y-vectors corresponding to the x-vectors belonging to the neighborhood of the current input vector $\mathbf{x}$. The weights for nearest neighbours can be all the same $(1/k)$ or can express similarity between vectors in X-space.

The forecasting model based on *k*-NN was proposed in [3]. In this work we expand this model by incorporation additional predictors corresponding to weather factors. This leads to generalization of the nearest neighbor model and can improve its accuracy.

Nearest neighbour forecasting model

A vector of predictors $\mathbf{x}$ is called an input pattern. Their components represent successive demand or load values from the daily period: $\mathbf{L}_i = [L_{i,1}\ L_{i,2}\ \dots\ L_{i,n}]$, where $L_{i,t}$ is the power system load at period t of the day i and n is typically 24, 48 or 96. The load vector $\mathbf{L}_i$ is mapped into pattern $\mathbf{x}_i = [x_{i,1}\ x_{i,2}\ \dots\ x_{i,n}]$ as follows:

$$x_{i,t} = \frac{L_{i,t} - \overline{L}_i}{\sqrt{\sum_{l=1}^{n}(L_{i,l} - \overline{L}_i)^2}} \tag{2}$$

where $\overline{L}_i$ is the mean load of the day i.

Equation (2) expresses normalization of vectors $\mathbf{L}_i$. After normalization they are unified: they all have unity length, zero mean and the same variance.

The response vector $\mathbf{y}$ is called an output or forecast pattern. The output pattern $\mathbf{y}_i = [y_{i,1}\ y_{i,2}\ \dots\ y_{i,n}]$ maps the load vector representing successive loads from the day $i+\tau$: $\mathbf{L}_{i+\tau} = [L_{i+\tau,1}\ L_{i+\tau,2}\ \dots\ L_{i+\tau,n}]$, where $\tau > 0$ is a forecast horizon:

$$y_{i,t} = \frac{L_{i+\tau,t} - \overline{L}_i}{\sqrt{\sum_{l=1}^{n}(L_{i,l} - \overline{L}_i)^2}} \tag{3}$$

The goal of mapping the response vectors into y-patterns is their unification. Unified x- and y-patterns express shapes of daily curves. Annual and weekly variations and also trend are filtered out. Note that in (3) we use the known current process parameters for the day i: $\overline{L}_i$ and $\sqrt{\sum_{l=1}^{n}(L_{i,l} - \overline{L}_i)^2}$, instead of parameters for the day $i+\tau$ which are unknown. This enables us to determine the forecast of vector $\mathbf{L}_{i+\tau}$ using transformed equation (2) after the forecast of pattern $\mathbf{y}$ is generated by the model.

The i-th input and output patterns are paired and included in the training set $\Omega = \{(\mathbf{x}_1, \mathbf{y}_1), ..., (\mathbf{x}_N, \mathbf{y}_N)\}$. For the current input pattern $\mathbf{x}$ its k nearest neighbours in the training set Ω are selected. Let us denote successive nearest x-patterns as: $\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^k$, and their corresponding y-patterns as: $\mathbf{y}^1, \mathbf{y}^2, ..., \mathbf{y}^k$. We call these k y-patterns the construction patterns, because the output y-pattern is estimated using (4) based only on these patterns:

$$\widehat{m}(\mathbf{x}) = \frac{\sum_{j=1}^{k} w(\mathbf{x}, \mathbf{x}^j)\mathbf{y}^j}{\sum_{j=1}^{k} w(\mathbf{x}, \mathbf{x}^j)} \quad (4)$$

The weighting function is of the form [4]:

$$w(\mathbf{x}, \mathbf{x}^j) = a\left(\frac{1 - \dfrac{d(\mathbf{x}, \mathbf{x}^j)}{d(\mathbf{x}, \mathbf{x}^k)}}{1 + b\dfrac{d(\mathbf{x}, \mathbf{x}^j)}{d(\mathbf{x}, \mathbf{x}^k)}} - 1\right) + 1 \quad (5)$$

where: a and b are parameters, and $d(\mathbf{x}, \mathbf{x}^j)$ is the distance between patterns $\mathbf{x}$ and $\mathbf{x}^j$.

Function (5) is shown in Fig. 1. Note that more distant neighbours have smaller weights and consequently less impact on shaping the model response.

Fig. (2) shows the *k*-NN forecasting model scheme, where $w'(\mathbf{x}, \mathbf{x}^j) = w(\mathbf{x}, \mathbf{x}^j) \Big/ \sum_{j=1}^{k} w(\mathbf{x}, \mathbf{x}^j)$ are the normalized weights satisfying the condition $\sum_{j=1}^{k} w'(\mathbf{x}, \mathbf{x}^j) = 1$.

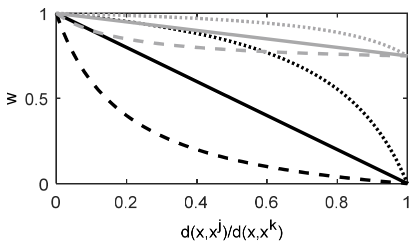

Fig.1. Weighting function (5) for b = 0 – continuous lines, b = 5 – dashed lines, b = –0,8 – dotted lines, a = 1 – dark lines and a = 0,25 – light lines.

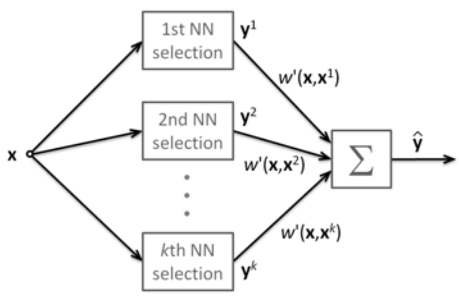

Fig.2. Scheme of the *k*-NN model.

Weather factors as forecast pattern contexts

To include weather factors as exogenous variables into the *k*-NN model we define contexts of the forecast pattern. Contexts for representing weather factors in pattern similarity-based forecasting models were proposed in [5]. By a context we mean a factor influencing the forecasted load, e.g. atmospheric temperature, humidity, wind speed, etc. Contexts related to different factors can be expressed by scalars or vectors. For example, mean daily temperature or minimal and maximal daily temperatures or series of 24 hourly temperatures in the day of forecast. The learning sample is extended to include m contexts: $(\mathbf{x}, \mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_m, \mathbf{y})$, where $\mathbf{c}_l$ is the l-th context vector or scalar. When we denote a set of predictors as $p = \{\mathbf{x}, \mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_m\}$, the learning sample is $(p, \mathbf{y})$.

In the forecasting phase, the k nearest neighbours for the current predictors p are selected in Ω, based on the combined distance taking into account x-patterns and contexts:

$$d(p, p_i) = v_0 d(\mathbf{x}, \mathbf{x}_i) + \sum_{l=1}^{m} v_l d(\mathbf{c}_l, \mathbf{c}_{l,i}) \quad (6)$$

where: $\mathbf{c}_l$ and $\mathbf{c}_{l,i}$ are vectors of the l-th contexts corresponding to patterns $\mathbf{x}$ and $\mathbf{x}_i$, v_0 is the weight assigned to the distance between x-patterns and $v_1, v_2, ..., v_m$ are weights assigned to distances between successive contexts, $\sum_{l=0}^{k} v_l = 1$.

The distances between x-patterns, $d(\mathbf{x}, \mathbf{x}_i)$, and contexts, $d(\mathbf{c}_l, \mathbf{c}_{l,i})$, should be from the same range to unify their impact on the combined distance (6). This impact can be controlled by weights v_l adjusted during training.

Let us denote successive nearest neighbours of p as: $p^1, p^2, ..., p^k$. The weighting function has the form (5) but instead of distances between x-patterns we use distances between predictors (6):

$$w(p, p^j) = a\left(\frac{1 - \dfrac{d(p, p^j)}{d(p, p^k)}}{1 + b\dfrac{d(p, p^j)}{d(p, p^k)}} - 1\right) + 1 \quad (7)$$

The regression *k*-NN model with contexts is of the form:

$$\widehat{m}(\mathbf{x}) = \frac{\sum_{j=1}^{k} w(p, p^j)\mathbf{y}^j}{\sum_{j=1}^{k} w(p, p^j)} \quad (8)$$

Note that the influence of the construction pattern $\mathbf{y}^j$ on the regression curve depends not only on the similarity between $\mathbf{x}^j$ and the current input pattern $\mathbf{x}$, as in (4), but also on the similarities between contexts of $\mathbf{y}^j$ and contexts of the forecasted pattern $\mathbf{y}$.

Application examples

The proposed *k*-NN model with exogenous variables was implemented in MATLAB and tested in one day ahead load forecasting (τ = 1) on several datasets, briefly described in Table 1. These data are available on the websites of system operators [6]. Utilized time series are intended to illustrate power systems of various size and demand characteristics. Each dataset includes hourly load time series (n = 24) and covers three years, from 2011 to 2014. Data from year 2011 to 2013 were used for model training and data from the last year were used for testing the model – created forecast was used as *ex-post* validation. Mean absolute percentage error (MAPE) was applied as a forecast quality measure in the following experiments.

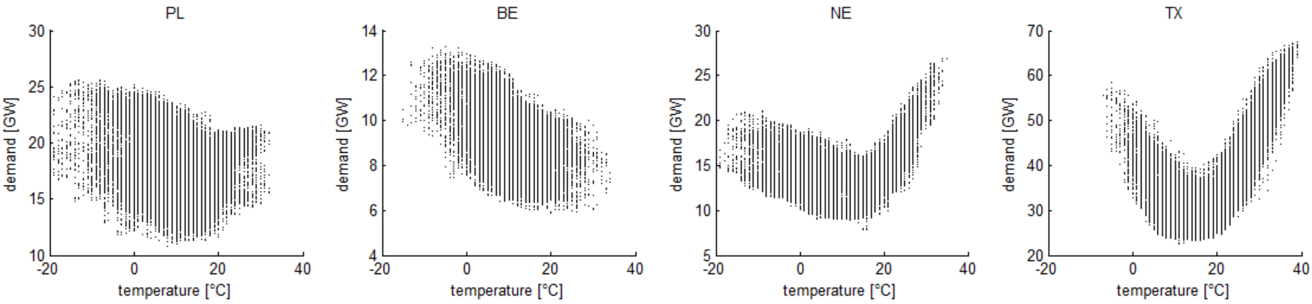

Fig.3. Temperature-demand relationship for datasets

Table 1. Datasets description

Label	Name	Mean demand
PL	PSE, Poland	18.2 GW
BE	Elia, Belgium	9.1 GW
NE	ISO-NE, New England	14.4 GW
TX	ERCOT, Texas	38.1 GW

Weather variables available for datasets are dry-bulb temperature and dew point temperature. Of these two, the former was chosen to define pattern context because of better results obtained during preliminary tests. Scatter plots showing temperature-demand relationship for our datasets are presented in Fig. 3. As we can see this relationship can be different depending on the power system features and geographical location. Strong nonlinearity of this relationship is observed for NE and TX. Such visible increase in electrical demand when temperature is decreasing below some threshold, as well as when it is increasing above it, has been already analysed in several papers [7, 8], and is mostly caused by HVAC (heating, ventilation, and air conditioning) equipment. Imbalance between heating and cooling components across different countries is attributed to different economic and climatic conditions [8].

Various versions of the proposed *k*-NN model was examined. All of them use the Euclidean metric as a distance measure. Differences between model variants concern the way of selection of neighbours and weather variables used as contexts. The model variants depending on the way of selection of neighbours are:

- WD – neighbours chosen from (**x**, **y**) pairs, where **y** is the same weekday as the forecasted day.
- DT – neighbours chosen from (**x**, **y**) pairs, where **y** is the same type of day (workday, Saturday, Sunday, holiday) as the forecasted day.
- WDT1 – if forecasted day is a holiday, neighbours chosen from (**x**, **y**) pairs, where **y** is a holiday too. Otherwise, neighbours are chosen as in WD.
- WDT2 – if forecasted day is a holiday, neighbours are chosen from (**x**, **y**) pairs, where **y** is the same weekday. Otherwise, neighbours are chosen as in DT.

The model variants depending on the weather extensions are:

- A – one context vector including 24 temperatures in each hour of the day represented by input pattern **x**.
- B – one context vector including 24 temperatures in each hour of the forecasted day represented by output pattern **y**.
- C – two context vectors, $\mathbf{c}_1$ and $\mathbf{c}_2$: $\mathbf{c}_1$ is same as used in A, $\mathbf{c}_2$ is same as used in B.

In B and C variants we use real temperatures for the day of forecast rather than predicted ones (ideal forecast). This let us to take weather forecast error aside from error introduced by our model.

In a first step of model building, variants without exogenous variables were optimized to find the best number of neighbours k and parameters of weighting function (5), a and b. With assumed ranges: k = 1,…, 20, a = 0, 0.1,…,1 and b = -0.99, 0.9,0.8,…,80, the best values of parameters were found using exhaustive search. The aim was to find for every model variant (WD, DT, WDT1, WDT2) such set of parameters that minimizes MAPE averaged over all types of days and datasets. The optimal values found were: k = 13 (WD, WDT1) and k = 14 (DT, WDT2), a = 1 and b = 20 for all models.

In a next step, the best values of weights v_l for combined distance (6) were found. Again exhaustive search was applied to check all valid combination of weights, assuming $v_l \in [0,1]$, $\sum_{l=0}^{k} v_l = 1$ and searching step of 0.01. Separate optimization procedures were executed: for every model variant, for every type of day (workday, weekend, holidays) and for every dataset. Motivation for that was to get better understanding of temperature influence on demand for different types of days. This step has demonstrated that:

- for the PL and BE datasets, the models gave best results when temperature was used in small extent (BE) or not used at all (PL).
- for the NE and TX datasets, situation was almost opposite. Depending on the model's variant, the best results for workdays were obtained when temperature weights were from 0.1 to even 0.5. The TX dataset is more temperature-dependent of these two. On the other hand, holidays demand seems to be quite insusceptible for temperature – most variants gave better results for temperature weights near to zero.

Finally, in the validation phase, the optimized models with exogenous variables were compared with their univariate versions (without exogenous variables) and with the naïve model. Simple naïve model using data from 7 days before the forecasted day as a prediction was used as a benchmark.

Results of final evaluation are shown in Table 2. Introduction of weather variables into *k*-NN model gave noticeable increase in forecast accuracy for the working and weekend days of NE and TX datasets (holidays have seen no improvement). Improvement for the PL and BE datasets was not clearly observed. Although accuracy of forecasts for weekends of BE datasets were improved, for other types of days accuracy was at the same level or even decreased. For working and weekend days of both NE and TX datasets, all variants of the model (WD, DT, WDT1, WDT2) benefited from introduction of temperature data. The best results for workdays were observed when two hourly temperature curves corresponding to the days represented by both x- and y-patterns are used as contexts (variant C). Worse results were obtained in variant B, where the context vector only consists of temperatures for the forecasted day. And the least impact on the forecast accuracy was observed when the context consists of temperatures for the day represented by x-pattern (variant A).

Table 2. MAPE of next day load forecasts created by models for all datasets.

Model	PL			BE			NE			TX		
	workday	weekend	holiday	workday	weekend	holiday	workday	weekend	holiday	workday	weekend	holiday
Naïve	4,02	4,11	24,61	5,18	4,73	18,51	6,59	7,08	7,14	11,88	10,55	10,38
WD	1,66	1,75	15,04	2,84	3,12	12,54	3,26	3,16	6,78	4,82	4,99	6,59
WD A	1,66	1,75	15,71	2,84	3,12	12,34	3,22	3,16	7,03	4,68	4,86	7,08
WD B	1,67	1,75	16,17	2,82	3,02	12,41	3,03	3,02	6,78	4,20	4,44	6,17
WD C	1,67	1,75	16,17	2,82	3,02	12,41	2,91	2,95	7,03	4,05	4,28	6,29
DT	1,56	1,76	4,39	2,87	3,12	4,10	3,08	3,16	6,47	4,46	4,98	7,62
DT A	1,55	1,76	4,45	2,87	3,12	4,26	3,01	3,15	6,47	4,35	4,84	7,62
DT B	1,55	1,76	4,44	2,88	3,03	4,24	2,67	3,02	6,47	3,68	4,44	7,62
DT C	1,55	1,76	4,44	2,88	3,03	4,24	2,41	2,92	6,47	3,43	4,31	7,62
WDT1	1,66	1,75	4,36	2,84	3,12	4,05	3,26	3,16	6,45	4,82	4,99	7,54
WDT1 A	1,66	1,75	4,39	2,84	3,12	4,19	3,22	3,16	6,45	4,68	4,86	7,61
WDT1 B	1,67	1,75	4,37	2,82	3,02	4,15	3,03	3,02	6,45	4,20	4,44	7,54
WDT1 C	1,67	1,75	4,38	2,82	3,02	4,15	2,91	2,95	6,45	4,05	4,28	7,61
WDT2	1,56	1,76	15,00	2,87	3,12	12,23	3,08	3,16	6,36	4,46	4,98	6,15
WDT2 A	1,55	1,76	15,15	2,87	3,12	12,61	3,01	3,15	6,35	4,35	4,84	6,90
WDT2 B	1,55	1,76	15,96	2,88	3,03	12,40	2,67	3,02	6,36	3,68	4,44	6,05
WDT2 C	1,55	1,76	15,54	2,88	3,03	12,40	2,41	2,92	6,36	3,43	4,31	6,20

The biggest improvement was obtained by models DT and WDT2. With weather extension of type C, MAPE of models DT and WDT2 for NE dataset was reduced from 3.08% to 2.41% (workdays) and from 3.16% to 2.92% (weekends). For TX dataset, reduction was from 4.46% to 3.43% (workdays) and from 4.99% to 4.28% (weekends). Simultaneously, accuracy of forecasts for holidays seem to obtain no benefits from introduction of exogenous variables. Some of the model variants gave a little better results after this introduction, but results of other ones worsened with no visible regularity.

Conclusions

In this paper we propose a *k*-nearest neighbour model with exogenous variables representing weather factors to a problem of electricity demand forecasting in short horizons. Weather factors are represented by contexts of forecast patterns. Contexts are taken into account in selection of the nearest neighbours.

In application examples extended model provides noteworthy improvement in forecasts for two of four considered datasets and it seems to be some pattern in this result. The scatter plots (Fig.1) demonstrate, that the power systems which are the source of data can be assigned to two groups. The first one is characterized by more linear temperature-demand relationship (PL and BE) and the second one is characterized by strong nonlinearity in this relationship, possibly due to much bigger participation of cooling component in the overall demand (NE and TX). An increase in the forecast accuracy was achieved only for the latter group. For such temperature-sensitive power systems, we can conclude that forecasting accuracy can be improved by temperature contexts. The biggest improvement occurred when contexts represented the temperatures for the forecasted day. But even in the case of using past temperatures (for the day corresponding to input x-pattern) as a context, improving was observed. In our study we used real temperatures for the forecasted day which in fact are not known in the moment of forecasting. In practice the forecasted temperatures are used. This leads to deterioration in the model quality caused by inaccuracy in temperature forecasting. So we can expect worse results for variants B and C than presented in Table 1.

Another difference between two groups of datasets is that PL and BE demonstrate noticeable differences between demand during holidays and regular days. Models trying to handle holidays in a regular manner, searching neighbours among days of the same weekday (WD and WDT2) were failing badly. Results for NE and TX exhibit opposite tendency. Holidays' forecasts created by WD and WDT2 models (looking for weekday rather than type of day) were not worse, but even a little better in many cases than those of models DT and WDT1.

***Authors**: Grzegorz Dudek PhD, Czestochowa University of Technology, Institute of Computer Science, al. Armii Krajowej 17, 42-200 Czestochowa, Poland, E-mail: Dudek@el.pcz.czest.pl; Marcin Janicki, PhD student, Czestochowa University of Technology, al. Armii Krajowej 17, 42-200 Czestochowa, Poland, E-mail: Marcin.janicki.poczta@gmail.com.*

REFERENCES

[1] Michie D., Spiegelhalter D.J., Taylor C.C.: Machine Learning, Neural and Statistical Classification, Ellis Horwood, (1994)

[2] Cichosz P.: Systemy uczące się, WNT (2007)

[3] Dudek G.: Prognozowanie krótkoterminowe obciążeń systemów elektroenergetycznych z wykorzystaniem estymatorów najbliższego sąsiedztwa, *Mat. konf. Aktualne Problemy w Elektroenergetyce*, t. 3 (2009), 31-39

[4] Dudek G.: Systemy uczące się oparte na podobieństwie obrazów do prognozowania szeregów czasowych obciążeń elektroenergetycznych. Akademicka Oficyna Wydawnicza EXIT, Warszawa (2012)

[5] Popławski T., Dudek G., Łyp J.: Forecasting Methods for Balancing Energy Market in Poland, *Int. Jour. Electrical Power and Energy Systems*, 65 (2015), 94-101

[6] PL (PSE): http://pse.pl/index.php?dzid=77,
BE (Elia): http://www.elia.be/en/grid-data/data-download,
NE (ISO-NE): http://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/zone-info,
TX (ERCOT): http://www.ercot.com/gridinfo,
Weather data (NCDC): http://www7.ncdc.noaa.gov/CDO/cdo

[7] Sailor D.J., Munoz J.R.: Sensitivity of electricity and natural gas consumption to climate in the USA - methodology and results for eight states, *Energy*, 22 (1997), no. 10, 987-998

[8] Bessec M., Fouquau J.: The non-linear link between electricity consumption and temperature: a threshold panel approach. *Energy Economics*, 30 (2008), no. 5, 2705-2721