Marcin JANICKI

Czestochowa University of Technology, Faculty of Electrical Engineering

doi:10.15199/48.2017.03.03

Temperature correction method for pattern similarity-based short-term electricity demand forecasting models

***Abstract**. In this paper, a temperature extension of pattern similarity-based (PSB) short-term load forecasting models is proposed. Different variants of these models were recently thoroughly described in literature, though focus was placed on univariate-type ones. Proposed method introduces correction of temperature bias into the model. PSB model with proposed correction is examined on several datasets illustrating power systems with various demand characteristics. Different variants of method are investigated to evaluate its influence on forecasting performance of the model.*

***Streszczenie.** W pracy przedstawiono temperaturowe rozszerzenie modeli prognostycznych opartych na podobieństwie obrazów. Proponowana metoda wprowadza do modeli korekcję uwzględniającą wpływ temperatury na zapotrzebowanie na energię. Działanie różnych wariantów przedstawionej korekcji jest badane na zbiorach danych pochodzących z różnych systemów energetycznych. (**Metoda korekcji temperaturowej modeli opartych na podobieństwie obrazów do prognozowania krótkoterminowego zapotrzebowania na energię elektryczną**).*

Keywords: short-term load forecasting, pattern similarity-based models, temperature correction
Słowa kluczowe: prognozowanie krótkoterminowe obciążeń, modele oparte na podobieństwie obrazów, korekcja temperaturowa

Introduction

Short-term load forecasting (STLF) is a problem of considerable significance for functioning of power systems operating in environment of deregulated energy market. Accurate forecasts are important for power systems operators for both technical and economic factors [1]. Different approaches aiming to solve this problem have been presented. Most of them can be classified into one of two groups [2, 3]: conventional (time series models, smoothing techniques, autoregressive moving average) and soft-computing (ANN, Fuzzy-Neural systems, Expert systems). Another group is formed by similarity-based methods. These methods are generalization of minimal distance methods and use similarities between time series sequences to estimate the regression function [4]. In the following chapter, a brief overview of pattern similarity-based models (PSB) described in literature is presented. Afterwards, method of improvement by temperature correction for one of such models is proposed. Finally, application examples of method applied to real demand and temperature data are provided and discussed.

Pattern similarity-based models

Electrical load of power system is an example of time series with periodicities. In general, daily, weekly and annual variations over time are visible there as well as trend component [3]. These series are non-stationary in mean and variance and in addition contain random noise. The future value of time series can be deduced from past value what took place in a similar conditions (similar weekday, type of day, weather conditions). However, elements of time series like weekly and annual seasonal variations and trend should be removed from time series. Patterns of load time series are created by preprocessing procedure using one of several function mappings. Pattern is a vector where every element was created by transforming corresponding element of time series with use of specified function [3]. Pairs of subsequent patterns are created in a way that time interval between preceding pattern $\mathbf{x}$ and following pattern $\mathbf{y}$ is constant value labelled as a forecast horizon [3]. The general forecasting method of PSB models is based on assumption formulated in [4, p. 163]: *"If the process pattern $\mathbf{x}_a$ in a period preceding the forecast moment is similar to the pattern $\mathbf{x}_b$ from the history of this process, then the forecast pattern $\mathbf{y}_a$ is similar to the forecast pattern $\mathbf{y}_b$"*. On the basis of this assumption, forecast $\mathbf{y}_b$ can be created using history data in a form of $\mathbf{x}_a$, $\mathbf{x}_b$ and $\mathbf{y}_b$. The only prerequisite is the existence of similarity measure among patterns.

In general PSB forecasting models implement procedure consisting creation of patterns, selection from them ones having preceding part ($\mathbf{x}$-pattern) similar to preceding part of forecasted time series and finally creating forecast with use of $\mathbf{y}$ patterns of selected subset [3].

Several variants of PSB models have been described in literature. Differences between them concern a way of pattern creation and method evaluating similarity between patterns. For similarity evaluation many methods including nearest neighbour estimators and kernel estimators of function regression have been described [5]. Other details concerning structure and functioning of these models are irrelevant in a context of this paper and will be not considered here.

Temperature influence on electricity demand

Electrical demand is influenced by many interdependent factors, including calendar (weekly and annual periodicities) and weather variables [2]. Among weather-related factors, temperature is reported to be the most influential [6]. In general, temperature influences demand in a non-linear way. It is visible that increase in electrical demand can be observed in two opposite situations: when temperature decrease below some threshold and when it is increasing above some value [7]. It is believed that main reason of such form of this relationship originates from usage of HVAC (heating, ventilation, and air conditioning) equipment [6]. Depending on economic situation, geographical location and climate, the proportion of heating and cooling component can be different.

The attempt to separate weather-related component from total electricity consumption has been made by Hobby and Tucci and was described in [6]. The authors proposed method to estimate how much of total electricity consumption is used for cooling and heating with use of human-perceived equivalent temperature calculated from temperature and humidity (apparent temperature). For every hour of the day, the electricity consumption is plotted as a function of the apparent temperature. Plotted data is interpolated with a cubic polynomial and shifted down in such a way that its minimum value is zero. Such curve is the model of weather-related electricity consumption for considered hour [6]. Hence, to completely estimate the value of this component throughout a day, 24 curves are necessary. Result of application of this procedure to one of the mentioned datasets [9] is presented on Figure 1.

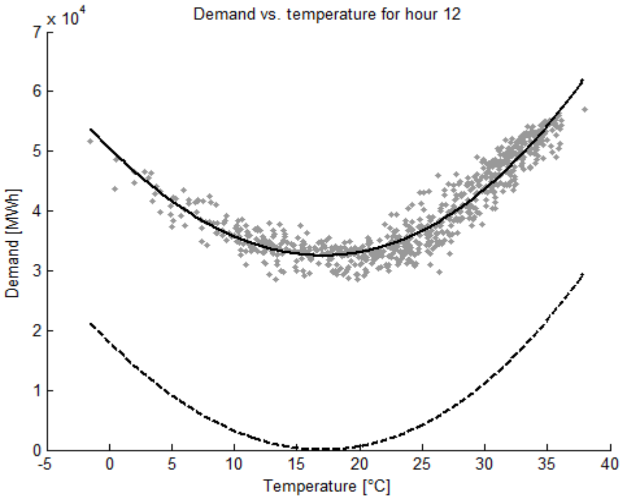

Figure 1. Example of weather-related component extraction [6] applied to one of datasets: interpolated curve (solid line) and shifted curve representing weather component (dashed line)

The reason of this procedure's introduction is justified by the assumption of existence of a temperature's comfort zone with no heating or cooling action necessary [6, Appendix A, p.6].

Temperature correction method

In reference to described research papers [6, 7] it is plausible to assume that past demand data is biased with temperature of the time. Past data contains usually whole spectrum of days, so temperatures for most data could be quite different than temperature expected on forecasted day. Because of this, patterns of demand time series created within described univariate similarity-based models are burdened with unadjusted temperature bias. Two possible remedies could be applied to overcome this obstacle:

- For every day being forecasted, past data used for creating forecast has to be restricted to days with similar temperature profile – patterns should carry also information about temperature.
- Temperature-related component has to be separated from demand to obtain temperature-independent historical data used to create patterns and then modify estimated forecast according to expected temperature.

The variant of the first solution is proposed in [1] where weather contexts for patterns similarity-based models are described. Extending regular demand data patterns with contexts causes taking weather conditions into account in similarity evaluation.

This paper explores a solution formulated on basis of the second solution. The following hypothesis is made: *forecasting accuracy of similarity-based models using univariate demand patterns can be improved by providing them data with weather conditions bias removed and moving process of including temperature influence into separated module*.

On this assumption, correction method is formed. Weather influence is limited to temperature only and primitive variable rather than its apparent derivative is used. Removal of temperature influence to obtain unbiased demand data is essential for the method. Modified data is then used as an input for selected pattern similarity-based model. Resulting output forecast of the model is devoid of temperature bias. As temperature influence on a demand on forecasted day has to be finally included, correction estimated with use of temperature predicted for this day is applied as the last step.

Complete procedure including temperature correction and building forecast by PSB model (as described in [4]) is as follows:

```
1. Evaluate   temperature-related   component   of
   demand with use of historical data.
2. Remove the temperature bias from past demand
   data using estimation from step 1.
3. Do regular forecasting by PSB model:
   3.1. Encode demand data modified in step 2 to
        patterns.
   3.2. For every forecasted day select similar x-
        patterns with use of similarity criterion.
   3.3. On a basis of patterns selected in step
        3.1, use an estimator to construct a y-
        pattern of forecast.
   3.4. Decode y-patterns created in step 3.3 to
        obtain a forecasted demand values.
4. For every built forecast use value of expected
   temperature to calculate temperature-related
   component and add it to obtain final value.
```

Algorithm 1. Forecasting procedure with temperature correction

Evaluation of the temperature-related component follows original procedure described in [6], although various modifications are introduced. Base version, defines energy consumption $F(h,T)$ as a sum of weather-related consumption C_{WR} and non-weather-related consumption C_{NWR} as [D, Appendix A]:

(1) $$F(h,T) = C_{WR}(h,T) + C_{NWR}(h,T)$$

where: h = 1,…,24 is the hour of day, T – temperature. Then for each hour h_0 and temperature T, expectation $E(F(h_0,T))$ of consumption function is:

(2) $$E(F(h_0,T)) = E(C_{WR}(h_0,T)) + E(C_{NWR}(h_0,T))$$

Non-weather-related component does not depend on T, so:

(3) $$E(F(h_0,T)) = E(C_{WR}(h_0,T)) + E(C_{NWR}(h_0))$$

The main assumption is the existence of comfort temperature $T_{min}(h_0)$ for hour h_0 when consumption due to HVAC is zero [6]:

(4) $$E(C_{WR}(h_0,T_{\min}(h_0))) = 0$$

Then, expectation of weather-related consumption can be approximated by polynomial $P(h_0,T)$ in T with coefficients depending on h_0:

(5) $$E(C_{WR}(h_0,T)) = P(h_0,T) - \min(P(h_0,T))$$

Analogously, reasoning described before is extended in this paper to cover energy consumption functions depending on one or more variables including: temperature, hour of day, type of day (workday, Saturday, Sunday, holiday), day of week, season of year. From now, let us replace „weather-related" term with „temperature-related" as we consider it as a function of primitive temperature. Also, let us denote:

- Polynomial in T with coefficients depending on discrete variables a_1, …, a_n as $P(T, a_1, …, a_n)$. Every combination of a variables has its own coefficients for polynomial.

- Expectation of temperature-related consumption as $E_{TR}(.)$,
  $E_{TR}(T, a_1, \ldots, a_n) = P(T, a_1, \ldots, a_n) - \min(P(T, a_1, \ldots, a_n))$
- Temperature-related consumption calculated by estimator Z as $E_{TR}{}^{Z}(.)$
- Number of polynomials created by estimator Z as $N(Z)$.

Considered types of estimators for temperature-related consumption are listed below:

- A – for all days and one polynomial $P(T)$ is created to estimate $E_{TR}{}^{A}(T)$, $N(A) = 1$.
- B – for every hour h one polynomial $P(T,h)$ is created to estimate $E_{TR}{}^{B}(T,h)$, $N(B) = 24$.
- C – for every day type dt (workday, Saturday, Sunday, holiday) creates one polynomial $P(T,dt)$ to estimate $E_{TR}{}^{C}(T,dt)$, $N(C) = 4$.
- D – for every hour h and for every day type dt one polynomial $P(T,h,dt)$ is created to estimate $E_{TR}{}^{D}(T,h,dt)$, $N(D) = 96$.
- E – for every weekday wd one polynomial $P(T,\mathrm{wd})$ is created to estimate $E_{TR}{}^{E}(T,wd)$, $N(E) = 7$.
- F – for every hour h and for every weekday wd one polynomial $P(T,h,wd)$ is created to estimate $E_{TR}{}^{F}(T,h,wd)$, $N(F) = 168$.
- G – for every hour h and for every weekday wdx one polynomial $P(T,h,wdx)$ is created to estimate $E_{TR}{}^{G}(T,h,wdx)$, $N(G) = 192$. Variable wdx differs from wd because it treats holidays separately.
- H – for every hour h and for every season of year s one polynomial $P(T,h,s)$ is created to estimate $E_{TR}{}^{H}(T,h,s)$, $N(H) = 4$. Samples are assigned to seasons depending on their month (s_1 = Dec, Jan, Feb; s_2 = Mar, Apr, May; s_3 = Jun, Jul, Aug; s_4 = Sep, Oct, Nov).

Removal of temperature bias with use of estimator type Z is done in a following way:

$$D_Z = D - E_{TR}{}^{Z}(.) \tag{6}$$

where: D_z – value corrected with use of estimator Z, D – original value of sample, $E_{TR}{}^{Z}$ – temperature-related component calculated by estimator Z. For calculation of $E_{TR}{}^{Z}$ component temperature data of the time is used and another variables depending on type of estimator.

Including a temperature bias in a forecast created by PSB model is done by adding $E_{TR}{}^{Z}$ component calculated with use of temperature expected for forecasted day and, if necessary, other variables used by estimator (equation (7)):

$$F_Z = F + E_{TR}{}^{Z}(.) \tag{7}$$

where: F_z – final demand forecast, F – forecast as built by PSB model, $E_{TR}{}^{Z}$ – temperature-related component calculated by estimator Z.

Application examples

The model with proposed correction was implemented in MATLAB and tested in one day ahead load forecasting on two datasets [9]. First dataset, labelled ERCOT is for power system of Texas with average demand value of 38.1 GW. Second dataset, labelled ELIA is for Belgium with average demand of 9.1 GW. As visible on Figure 2, demand response for temperature is almost linear for ELIA dataset and highly nonlinear for ERCOT.

Each dataset includes hourly load time series and dry-bulb temperature. Data from years 2013 and 2014 were used as ex-post validation. While validating with given year, two preceding years were used as historical data for the model. Mean absolute percentage error (MAPE) is applied as a forecast quality measure in the following experiments. Results of forecasting with different types of correction estimators were compared with results obtained with model without correction and naïve model.

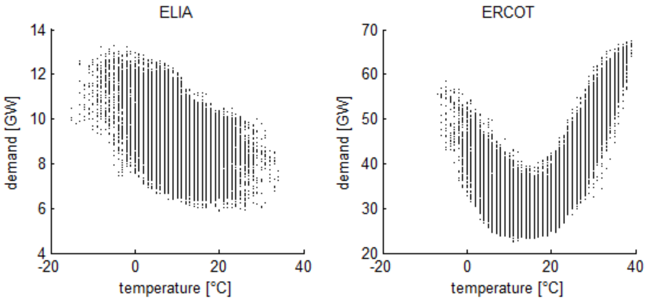

Figure 2. Temperature-demand relationships for datasets

PSB model enhanced with correction was k-Nearest Neighbour model, described in [8]. Input patterns $\mathbf{x}$ are defined as [4]:

$$x_{i,t} = \frac{L_{i,t} - \overline{L}_i}{\sqrt{\sum_{l=1}^{n} (L_{i,l} - \overline{L}_i)^2}} \tag{8}$$

where $L_{i,t}$ is the power system load at period t of the day i, n is 24 and $\overline{L}_i$ is the mean load of the day i. Output (forecast) patterns $\mathbf{y}$ representing successive loads from the day $i+\tau$: $\mathbf{L}_{i+\tau} = [L_{i+\tau,1}\ L_{i+\tau,2} \ldots L_{i+\tau,n}]$ are defined similarly as [4]:

$$y_{i,t} = \frac{L_{i+\tau,t} - \overline{L}_i}{\sqrt{\sum_{l=1}^{n} (L_{i,l} - \overline{L}_i)^2}}\,. \tag{9}$$

where $\tau > 0$ is a forecast horizon.

k-NN estimator of a model uses the Euclidean metric as a distance measure. Selection of neighbours is done in such a way, that search space is limited for patterns of days having the same type as forecasted day (workday, Saturday, Sunday, holiday). Number of neighbours $k = 14$. Weighting function for estimator was selected as described in [5, equation 16], parameters: $p = 1$, $\gamma = 20$.

All polynomials created to evaluate temperature-related component of demand were of maximum 3rd degree and gave best fit in a least-squares sense.

As a temperature expected for the forecasted day we use real values rather than predicted ones (ideal forecast). This let us to take weather forecast error aside from error introduced by our model.

Results of models examination are shown in Table 2. MAPE was calculated separately for workdays, weekdays and holidays.

With regular k-NN model as a reference, we see that improvement of accuracy is clearly visible for ERCOT dataset. For workdays and holidays, every variant of correction gave better results for both years 2013 and 2014. For weekend day there was also an improvement, although not unanimous, as variants E and G gave worse results than model without correction. Correction variant B (separate curve for every hour) performed best for workdays and weekdays of both years. For holidays there was no clear regularity and variants ranked differently for years 2013 and 2014.

Table 1. MAPE of forecasts created with use of examined models for year 2013 and 2014.

Model	dataset ERCOT						dataset ELIA					
	2013			2014			2013			2014		
	work	week	hol	work	week	hol	work	week	hol	work	week	hol
Naïve	10,31	10,19	11,84	11,88	10,55	10,38	5,79	4,67	15,16	5,18	4,73	18,51
PSB k-NN	3,92	4,38	7,63	4,46	4,98	7,62	**2,99**	3,21	5,75	**2,87**	3,12	4,10
PSB k-NN A	3,59	4,06	5,55	3,89	4,33	6,00	3,08	3,11	4,31	2,93	**3,05**	4,05
PSB k-NN B	**3,33**	**3,57**	5,88	**3,76**	**3,82**	6,41	3,22	**3,10**	**3,72**	3,02	3,20	4,84
PSB k-NN C	3,60	3,90	**5,06**	3,91	4,28	5,89	3,23	3,27	4,21	2,96	3,17	**3,56**
PSB k-NN D	3,42	3,84	5,45	3,86	3,90	6,51	3,37	3,15	4,40	3,12	3,24	4,93
PSB k-NN E	3,73	4,72	5,67	4,21	4,39	**5,60**	3,50	3,66	4,91	3,46	4,80	4,44
PSB k-NN F	3,51	3,77	5,64	3,99	3,91	6,31	3,84	3,25	5,81	3,20	3,24	4,97
PSB k-NN G	3,69	4,71	5,92	4,21	4,41	5,62	3,33	4,22	4,47	3,34	4,75	4,29
PSB k-NN H	3,50	3,97	5,56	4,04	4,54	6,26	3,30	3,42	4,79	2,93	3,04	4,07

On the other hand, for dataset ELIA, models with correction failed to provide and improvement. For workdays of both years all variants performed worse than regular k-NN model. For weekends and holidays, some variants were able to beat model without correction, although scale of improvement and lack of its regularity suggest rather incidental character of it.

Reason of different performance of correction method for both datasets seems to be caused by their different temperature-demand relationships (Fig. 2). ERCOT power system has non-linear demand response for temperature with clearly visible cooling component of the consumption. On the contrary, temperature-demand relationship of ELIA is almost linear and demonstrates no influence of cooling demand. This could be accounted to climate difference and higher average temperature throughout the year in Texas.

Results of different variants of the method show that there is no gain in making it more sophisticated. Good performances of simpler versions (B, C, D) were not improved when mixed together into more complex ones (D, F, G).

Conclusion

In this paper a temperature extension of pattern similarity-based (PSB) short-term load forecasting models is proposed. It introduces correction of temperature bias into model to take into account influence of weather on electrical demand

In application examples different variants of method were investigated to evaluate their contribution into overall forecasting performance of the model. Experiments carried out on two different datasets showed that possible improvement introduced by proposed correction method could depend substantially on temperature-demand characteristics of the power system. Performance of PSB model for system with non-linear response was improved for all types of days with use of proposed correction. Same model examined on system with linear response and lower average temperature obtained no benefit. We have observed general decrease of accuracy for workdays, sporadic increase for weekdays and holidays.

Investigations on different variants of the method show that in general simpler variants using smaller number of curves, but interpolated with use of bigger number of data points overperformed more complicated versions.

In our study we used real temperatures for the forecasted day instead of forecasted ones which are used in practice. For forecasted temperatures we can expect higher forecast errors.

Author: *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] Popławski T., Dudek G., Łyp J., Forecasting Methods for Balancing Energy Market in Poland, *International Journal of Electrical Power and Energy Systems*, 65 (2015), 94-101
[2] Kyriakides E., Polycarpou M., Short Term Electric Load Forecasting: A Tutorial, *Studies in Computational Intelligence*, 35 (2007), 391-418
[3] Dudek G., Pattern Similarity-based Methods for Short-term Load Forecasting – Part 1 – Principles, *Applied Soft Computing*, 37 (2015), 277-287
[4] Dudek G., Similarity-based approaches to short-term load forecasting, In: *Forecasting models: methods and applications*. iConcept Press, 2010a, 161-78
[5] Dudek G., Pattern Similarity-based Methods for Short-term Load Forecasting – Part 2 – Models, *Applied Soft Computing*, 36 (2015), 422-441
[6] Hobby J.D., Tucci G.H., Analysis of the Residential, Commercial and Industrial Electricity Consumption, *IEEE PES Innovative Smart Grid Technologies Asia (ISGT)*, 1 (2011), 1-7
[7] Henley A., Peirson J., Non-linearities in Electricity Demand and Temperature: Parametric versus Non-parametric Methods, *Oxford Bulletin for Economics and Statistics*, 59 (1997), 1149–1162
[8] Dudek G., Systemy uczące się oparte na podobieństwie obrazów do prognozowania szeregów czasowych obciążeń elektroenergetycznych, *Akademicka Oficyna Wydawnicza Exit*, (2012), 125-129
[9] ELIA: http://www.elia.be/en/grid-data/data-download,
ERCOT: http://www.ercot.com/gridinfo,
Weather data (NCDC): http://www7.ncdc.noaa.gov/CDO/cdo.