Short-term forecasting of district heating demand

Roman Petrichenko, Dmitry Sobolevsky, Antans Sauhats

Faculty of Power and Electrical Engineering
Riga Technical University
Riga, Latvia
romans.petricenko@rtu.lv, dmitrijs.sobolevskis@rtu.lv, sauhatas@eef.rtu.lv

Abstract— **Focus of the paper is statistical data pre-processing before it applies for prediction the thermal load in district heating networks, focusing on day-ahead hourly planning. Such a planning is highly important for cogeneration plants participating in electricity wholesale markets. Article considers the possibility of correcting detected inconsistencies into district heating statistical data using forecasted values of the heat demand. The case study is based on the examples of heat supply of a large city, gas fired cogeneration power plants and real world data. The cost of errors in the prediction of heat consumption is estimated.**

Keywords — energy market, district heating, cogeneration, forecasting, error analysis.

I. INTRODUCTION

In the electric power industry management of the European Union (EU) countries a significant modernization procedure was implemented, which is based on the use of a liberalized energy market. As a result, competition phenomenon had arisen – the main goal of the producers and retailers is profit maximization under the rules of electricity market. Its structure comprises follows electricity market types:

- Day-ahead market.
- Intraday market.
- Balancing market [1].

The largest volume of energy is distributed based on the use of the first type of market. To participate in the day-ahead market, a generating company submits its bid based on the day-ahead operational planning. Once the power production bid is placed on the electricity market, it can be changed only with surplus payment.

In order to satisfy rules of the first energy market type, it is necessary for generation companies to have adequate production plan. Short-term forecasting of generated power substantially solves the design-making process of power system operating mode.

The task of short-term forecasting is particularly relevant for planning the operating mode of cogeneration power plants (CHPP). The significance of such plants is increasing, as the Energy Efficiency Directive requires that each EU country aiming at reducing air emissions and enhancing fuel efficiency uses the national capacity of cogeneration of district heating [2].

Cogeneration mode of CHPP is characterized by high dependence between volume of electricity production and the heat energy consumption. Therefore, the prediction of the heat demand, which, in turn, heavy reliance on the ambient temperature, serves as a basis for planning the electricity generation. Several techniques to forecast electrical load generation or heat load demand in a cogeneration mode of CHPP have been reported in the technical literature: artificial neural network (ANN), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), exponentially weighted moving average (EWMA), polynomial, fuzzy neural networks (FNN), weighted nearest neighbours (WNB), adaptive wavelet neural network (AWNN), hybrid intelligent system (HIS) [3-12].

The analysis of the above models and methods of forecasting the demand for thermal energy and electricity production allows us to conclude that they are based on the use of historical data on heat consumption.

Unfortunately, real-world data is often incomplete, inconsistent, and/or lacking in certain behaviours or trends, and is likely to contain many errors. Data pre-processing is a proven method of resolving such issues. Commonly known procedures of data pre-processing are: data cleaning, integration, transformation, reduction and discretization. Data cleaning procedure means filling of missed data, smoothing the noisy data, or resolving the inconsistencies [13-15]. Deletion of faulty data before appliance in heat load forecasting has been described in [4]. The authors of [11] and [12] conclude that analysis and respective correction during data pre-processing has high impact on toward forecast accuracy and, therefore, success or failure of a power system` profit-maximizing.

The main contribution of present work is improving the accuracy of the heat consumption forecast of a district heating system by identifying incomplete or erroneous historical heat demand data and correcting them through predicted values. The proofing of efficiency of the proposed statistical data pre-processing algorithm is based on the real-world data.

The remainder of this article contains the following sections: the prediction algorithm of heat load is discussed in Section II. The implications of faulty input data on the heat

demand forecasting in a district heating system as well as suggested pre-process procedure of historical data and comparison with several widely-known pre-processing methods are reviewed in Section III. Section IV reviews main results. Economic benefits of data pre-processing described in Section V and finally Section VI concludes the article.

II. THE PREDICTION ALGORITHM

Time series forecasting of heat and electricity production of CHPP is an extremely important asset. It is used daily and hourly to forecast energy usage and optimize scheduling and unit commitment. In order to forecast the value of thermal and electrical energy production, there are many techniques. As example, hourly records on the global heat and electricity production of CHPP can be applied.

Forecasting problem can be divided into several sub issues, namely ambient temperature prediction, thermal energy consumption forecast, production of electricity forecast in cogeneration mode. Based on previous studies, it can be surely asserted that the forecasting of the consumption of thermal energy is carried out by choosing the approximating procedure F, given by the parameters P and describing the demand for thermal energy Q at hour t at ambient temperature T in the same hour. The average absolute percentage error in the forecast of demand for heat load can be applied as a criterion in the procedure for detecting erroneous data. Approximating procedure F choice are based on the historical data records of heat demand $Q_{(th)}$ and ambient temperature $T_{(th)}$ [16]. The way the procedure is chosen depends on the approach used. So, for example, using artificial neural networks we choose: the type of network, the number of neurons and the set of coefficients. When using approximating polynomials, we choose the order of the polynomial and the coefficients. After determining the procedure F and its parameters P, we estimate the predicted demand for thermal energy $Q_{(tf)}$. In doing so, we use predictions of the ambient temperature $T_{(tf)}$.

$$Q_{(t)} = F(T_{(t)}) \qquad (1)$$

Block diagram in Fig.1 represents heat demand forecasting process. Necessary inputs of short-term heat demand forecast can be divided into internal and external data set. Historical records of heat demand and ambient temperature represent internal data set for short-term heat demand forecast, whereas ambient temperature prediction applies as external input data and is vital one. Weather forecasting has developed during thousand years and has reached high level in accuracy[17]. Despite on the technology development and stunning computational capabilities, the stochastic character of meteorological processes, that have significant effect on the forecast accuracy, in turn there are high chances that predictions of weather introduces the uncertainties into heat demand forecasting model [17].

III. FAULTY INPUT DATA EFFECT ON THE FORECASTING PROCEDURE

Lack in measured data of technological process, their erroneous values as well as due to the uncertainties that are present in the external processes implemented in forecasting, leads to fall of its accuracy [4], [11].

Historical thermal demand data must be pre-processed due to next causes:

- Rare occurrence of events that negatively effect on the forecast accuracy (broken pipes, major rebuilds, service work etc.). Under the same external conditions, historical data differ significantly.

- Incorrect filling of the database. Comprehensive study of historical data should reduce human factor of this tendency.

Fig. 2 shows thermal load historical data (red curve on left plane), respective experience in forecasting (blue curve on the left plane), actual heat demand (red curve on right plane) and its two predictions (orange and green dashed curves on the right plane). It can be seen a big drop in historical data (red curve) which as result leads to impossible forecasting explicitly with sufficient accuracy (orange dashed curve).

Meaningful improve of forecasting accuracy through correction of the faulty data via predicted values, is displayed by green dashed curve in Fig. 2.

The scope of this paper is accuracy increasing of heat demand forecast by means of recorded data pre-processing. Fig. 1 depicts pre-process block integration into the simple flowchart of heat demand forecasting.

Faulty data detection is realized by percentage calculation of distinction between actual heat demand value and its prediction value:

$$Deviation_t = \frac{|Q_{real(t)} - Q_{prediction(t)}| * 100}{Q_{real(t)}}, \qquad (2)$$

where $Deviation_t$ – percentage deviation at t-th hour (%), $Q_{prediction(t)}$ – heat demand forecast at t-th hour (MW), $Q_{real(t)}$ – actual value of heating demand at t-th hour (MW).

Analysis results of historical thermal load demand applying suggested faulty data detection method are displayed in Fig. 3. Bar chart depicts density of the detected possible incorrect data depending on the percentage deviation setting range. Analysing of historical heat demand has done for time range: $31(8^{00}).12.2014 – 31(7^{00}).03.2015 – 15(8^{00}).10.2015 – 21(7^{00}).04.2016 – 15(8^{00}).10.2016 – 28(8^{00}).03.2017$ (approximately accords to three heating seasons).
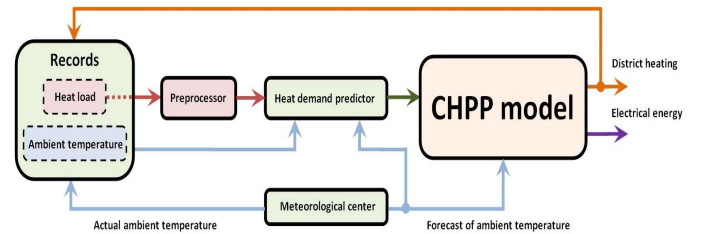

Fig. 1. Simplified flowchart of suggested preprocessing implementation into the foricasting heat demand
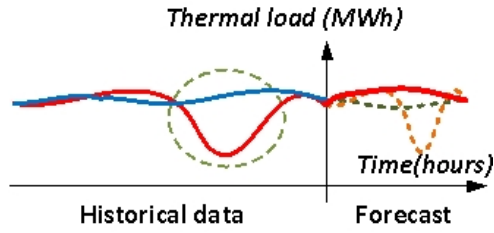
Fig. 2. Historical heat demand (red curve) and its forecast (blue curve). Heat demand prediction based on the corrected historical data (green dashed curve) and not preprocessed historical data (orange dashed curve). Bold red curve represents actual data of heat demand.

Dependence of the mean absolute percentage error (MAPE) of heat demand prediction inaccuracy on the historical time range is shown on Fig.4. After 60-th day the rise of the prediction inaccuracy is observed. MAPE average value of this period equals to the 4.82% and can be applied for magnitude calculation of deviation between actual and forecasted historical heat demand. Apart from that, such historical heat demand analysis gives possibility to mark detected faults dependently on the deviation magnitude (see Table 1).
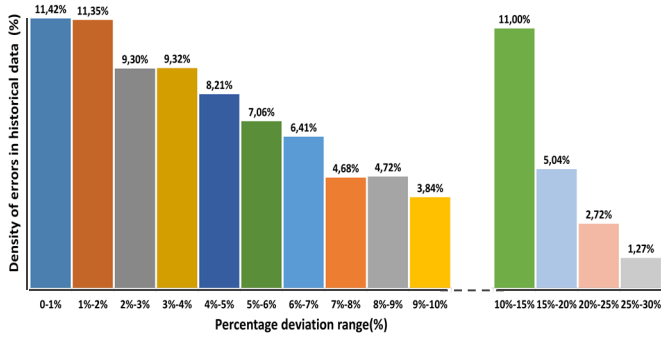

Fig. 3. Density of faults in thermal load historical data depending on the percentage deviation.
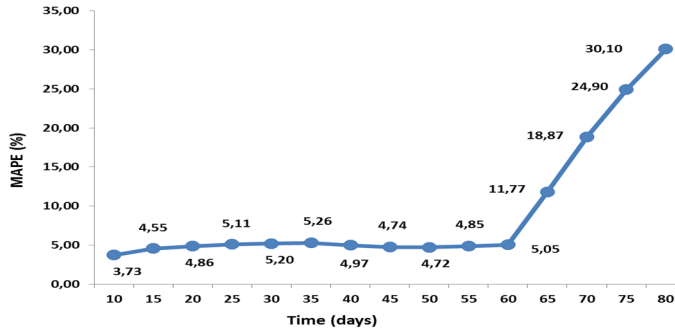

Fig. 4. Heat demand prediction inaccuracy dependence on the time range

TABLE I. COLOURED INTERPRETATION OF POSSIBLE FAULTS IN HISTORICAL HEAT DEMAND DATA

Probability of fault	Colour	Value of deviation
Low	Yellow	MAPE ≤ Deviation^p_t < 1.5* MAPE
Medium	Rose	1.5* MAPE ≤ Deviation^p_t < 2.0* MAPE
Hight	Orange	2.0* MAPE ≤ Deviation^p_t < 2.5* MAPE
Increased	Red	2.5* MAPE ≤ Deviation^p_t

IV. CASE STUDY

A. District heating system

Capacity of centralized district heat of city under consideration is more than half of whole country and equals to 1037 MWth. Two large cogeneration plants – CHPP-1 and CHPP-2 with electrical capacity 144 MWel, 832 MWel and thermal capacity 493 MWth, 544 MWth respectively – supplies 69% of the required heat demand. Total length of city`s heating network is about 800 km [18].

Aforementioned CHPPs mostly operate in the highly efficient cogeneration mode (generation of heat and power). The condensing and combined mode of operation is in most cases uncompetitive in the electricity market. Successful engagement of CHPP in day-ahead electricity market requires accurate short-term forecast of heat demand.

B. Analysis of heat demand historical data

Example of faulty historical heat demand data are shown in Tab.2, where: AT - ambient temperature (Celsius), *HD* – heat demand (MWh). At January 30 2017 6.^{00} the erroneous heat demand value is detected and marked correspondingly to the magnitude of the probability.

The proposed pre-process procedure of historical heat demand data highlights erroneous data by colour according to the probability level of faulty data. The forecast of heat demand, applying our previously developed polynomial model, is represented in Fig.5 [16].

TABLE II. HEAT DEMAND OF 29-30 JANUARY 2017.

| YEAR	MONTH	DATE	WEEK DAY	AT (C°)	HOUR	ACTUAL HD (MWh)	HD PREDICTION (MWh)
2017	1	29	7	-0,30	8	446	464,82
2017	1	29	7	-0,30	9	452	471,50
2017	1	29	7	-0,10	10	461	467,41
2017	1	29	7	0,00	11	470	464,59
2017	1	29	7	0,10	12	508	463,87
2017	1	29	7	0,10	13	519	461,51
2017	1	29	7	0,20	14	520	459,76
2017	1	29	7	0,20	15	520	464,91
2017	1	29	7	0,10	16	518	465,85
2017	1	29	7	0,10	17	518	468,69
2017	1	29	7	0,10	18	519	475,44
2017	1	29	7	0,10	19	518	478,89
2017	1	29	7	0,00	20	519	480,90
2017	1	29	7	-0,20	21	520	487,13
2017	1	29	7	-0,30	22	508	484,37
2017	1	29	7	-0,20	23	495	471,08
2017	1	30	1	-0,30	0	487	462,26
2017	1	30	1	-0,30	1	477	457,10
2017	1	30	1	-0,40	2	475	457,66
2017	1	30	1	-0,40	3	474	453,35
2017	1	30	1	-0,40	4	472	450,94
2017	1	30	1	-0,50	5	472	454,26
2017	1	30	1	-0,50	6	100	456,66
2017	1	30	1	-0,60	7	484	466,90
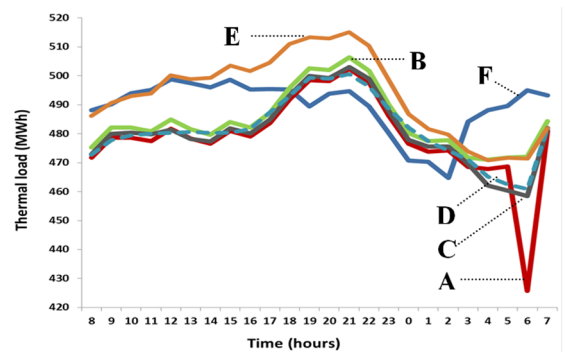
Fig. 5. Actual ('F' curve) and predicted values of heat demand on January 30-31 2017.

Diagram shows a comparison between the widely-known pre-process procedures and proposed method of data preparing further forecasting:

- Warnings ignoring (curve 'A'). Prediction of heat demand has been realized based on the non-corrected historical data - red solid curve and red bar in Fig.5, Fig.6 and Fig.7, accordingly. This action corresponds to 'conventional' heat demand prediction method [16].

- Prediction of heat demand after manual righting of detected erroneous data. Green solid curve ('B' behaviour) and green bar in Fig.5, Fig.6 and Fig.7.

- Filtration of last 24 hours of historical heat demand data using standard and weighted moving average. Black solid curve ('C' curve) and black bar in Fig.5, Fig.6 and Fig.7.

- Historical data filtration for 60 days back using standard and weighted moving average (corresponds to used average MAPE time range). Dashed cyan curve (curve 'D') and cyan bar in Fig.5, Fig.6 and Fig.7.

- Automatic correction of detected faulty data by means of the heat demand predicted value (curve 'E'). Solid orange curve and orange bar in Fig.5, Fig.6 and Fig.7.

Accuracies of the heat demand hour and day-ahead predictions are shown in Fig.7 and Fig.8, respectively.

With naive expectation that heat demand value is equal to previous one, green bar in 6th histogram represents highest accuracy of forecasted 6th hour heat demand. Suggested correction method of faulty data through the predicted values gives up leading position. Nevertheless, the forecast accuracy increase is above 9% (comparing with conventional technique of the thermal load forecasting). Fig. 7 depicts daily forecast precision. From histogram is clear, that the highest accuracy of the heat demand forecast is achieved by automatic correction of the detected faulty historical data – 97.67%.
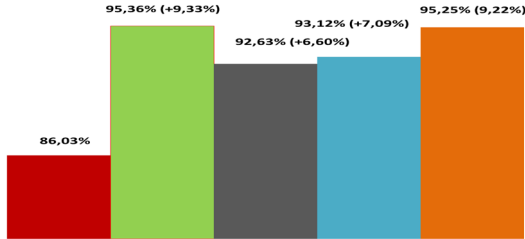

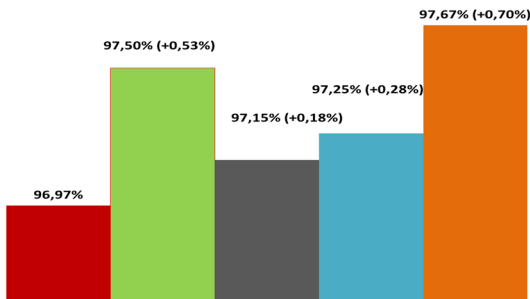Fig. 6. The 6th hour accuracies of heat demand predictions in30-31 January 2017.

Fig. 7. Accuracy of heat demand predictions in30-31 January 2017.

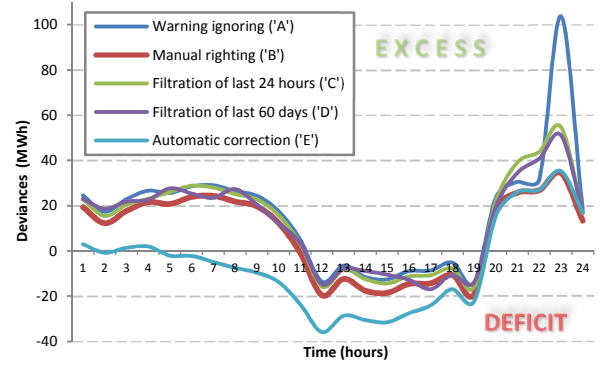Fig. 8. Inaccuracies in heat demand forecast for 30-31 January 2017

Increase of forecasting accuracy about 1% could yield savings in operational costs of CHPP as well as of power system as a whole. In monetary terms improvement of forecasting precision leads to earn of six-figure benefits [19].

V. ESTIMATION OF INACCURACIES COST

Curves of heat demand forecast, depicted in Fig.5, can be applied for the cost estimation of thermal load prediction inaccuracy. Fig. 8 represents the difference (Δdif) of the actual and predicted hourly values; this difference can be both, positive and negative.

The shortage of thermal energy, as well as its excess, leads to financial losses in the electricity market, since the operator of the CHPP is forced to buy or sell electric energy on the intraday market or to give up the cogeneration regime. To estimate the costs associated with negative or positive deviances from actual heat demand, some assumptions have to be done. We assume that the error in the forecast of heat consumption ΔQ causes an error in the forecast of the generated electricity ΔE. We accept, additionally, that the relationship between these errors is linear:

$$\Delta E = \Delta Q \cdot \alpha \qquad (3)$$

The alpha factor shows the ratio between the amount of electricity and heat generated. The CHPP under consideration are based on the use of high-efficiency gas turbines, providing an alpha value of about 1.5 [18].

Finally, we assume that if there is a need to buy energy, the market will have to pay a double price. In the case of its forced sale, the price will be only 30% of the market price of the first type. As a result, the cost of the error is estimated based on the expression:

$$\begin{cases} E = \sum_{i=1}^{n} |\Delta Q_i \cdot \alpha| \cdot \beta_i \cdot C_i \\ \Delta_{dif_i} < 0 \ \rightarrow \beta_i = 2 \\ \Delta_{dif_i} = 0 \ \rightarrow \beta_i = 1 \\ \Delta_{dif_i} > 0 \ \rightarrow \beta_i = 0.3 \end{cases} \qquad (4)$$

where E – cost of the forecast inaccuracy, ΔQ_i – thermal load value at i-th hour, α – ratio between the amount of electricity and heat generated, C_i – electricity market price at i-

th hour, n – time horizon, βi – electricity price coefficient, depending on the sign of the difference between actual and forecasted thermal load at i-th hour.

Using system of equation (4) and the previously mentioned data from Fig. 8 cost-benefit analysis of district heating forecast can be done. Table III lists data for some particular hours, identified in column 'T'. The historical heating demand is summarized in column 'HHD'. Heat demand forecast has obtained by the conventional method in the column labeled 'CON'. The 'DEV' column displays forecast errors in absolute values. To get electrical energy it multiplied by a coefficient α equal to 1.5 (p.u.). Electrical energy prices are shown in the column 'EP' (€/MWh). Column labeled 'S' shows the positive or negative status of the predicted thermal load in regard to the actual.

The corrected electricity prices (column 'CP' in €/MWh) depend from equation (3) and the direction of the forecast error 'S'.

Thus, it is possible to calculate the expenses caused by forecast inaccuracies (column 'EX' in €). The bottom-right cell in Table III shows the sum of hourly estimated expenses incurred due to imperfect forecast. For the selected day (January 30–31 2017), the total expense estimate reaches 35 600 €. Of course, this is an indicative number as the actual losses depend from the assumptions of β coefficient values and the real-time decisions made by the CHPP operators. The same cost estimating technology can be applied to the forecasts obtained after historical data mentioned preprocessing methods.

Table IV summarizes the results of thermal load forecast inaccuracy cost. Conventional method shows the higher forecast inaccuracy – 3.03%. Value of the thermal load prediction inaccuracy equals to the 35 600 €.

Proposed historical data preprocessing and the subsequent heat demand forecast, illustrates lowest prediction error – 2.33%. Thermal load prediction inaccuracy cost in this case is equals to the 11 207 €.

TABLE III.   INACURACY COST OF HEAT DEMAND FORECAST BY CONVENTIONAL METHOD

T	HHD	CON	DEV	EP	S	CP	EX
8	488.11	471.80	- 24.47	31.03	>	62.06	1 518.57
9	490.24	478.68	- 17.34	42.07	>	84.14	1 458.66
…	…	…	…	…	…	…	…
6	495.95	425.82	103.69	29.40	>	58.80	6 096.88
7	493.20	480.99	18.32	30.20	<	60.40	1 106.34
							Sum
							35 600 (€)

TABLE IV.   HEAT DEMAND FORECAST ERRORS AND THE CORRESPONDING PREDICTION INACURACCY COSTS

Date	'A'	'B'	'C'	'D'	'E'
Jan. 30–31, 2017	Mean absolute percentage error (MAPE) (%)				
	3.03	2.49	2.85	2.75	2.33
	Prediction inacuracy cost (€)				
	35 600	26 076	32 445	31 158	11 207

In addition, we note that the income from the sale of electricity on the day was EUR. The cost of errors is only a few percent, nevertheless, the absolute value of this loses is a significant amount.

The algorithm for preliminary data processing described in this article is implemented in the Matlab environment and is used in planning the operating modes of CHPP in Riga

VI.   CONCLUSION

Mean absolute percentage error of the heat load demand forecasts can be applied as criterion in faulty data detection procedure. A real-world data of single day heat load was considered in demonstrating the enhancement of forecast accuracy by suggested pre-processing algorithm. Reached heat demand prediction improvement allows saving on significant amount of energy sources increases the competitiveness and thereby expands substantially profit of power system. Financial analysis overviewed in article shows financial expenses due to prediction errors. Proposed method of historical data preprocessing increase the accuracy of forecasting tools, and, respectively, reduce economical loses during energy system exploitation process.

REFERENCES

[1] European power market 'Nord Pool': https://www.nordpoolgroup.com/the-power-market/

[2] European Commission Energy Efficiency Directive 2012/27: https://ec.europa.eu/energy/en/topics/energy-efficiency/cogeneration-heat-and-power

[3] E. Dotzauer, "Simple model for prediction of loads in district-heating systems," *ELSEVIER journal*, Applied energy vol. 73, pp. 277 - 284, 2002.

[4] N. Eriksson, "Predicting demand in district heating systems. A neural network approach," report of Upsala university, Maj 2012: http://www.diva-portal.org/smash/get/diva2:530099/FULLTEXT01.pdf

[5] H. Gadd, S. Werner, "Daily heat load variations in Swedish district heating systems," *ELSEVIER journal*, Applied energy vol. 106, pp. 47-55, 2013.

[6] S. Grosswindhager, A. Voigt, M. Kozek, "Online Short-Term Forecast of System Heat Load in District Heating Networks," in proceedings of 31st International Symposium on Forecasting, Prague, Czech Republic, 26-29 Jun 2011: https://publik.tuwien.ac.at/files/PubDat_202018.pdf

[7] N. Amjady, ''Day-ahead price forecasting of electricity markets by a new fuzzy neural network,' IEEE Trans. Power Syst., vol 21, no 2, pp. 887 – 896, May 2006.

[8] A. T. Lora, J. M. R. Santos, A.G. Exposito, J.L.M. Ramos, and J.C.R. Santos, "Electricity market price forecasting based on weighted nearest neighbors tehniques," IEEE Trans. Power Syst., vol 22, no 3, pp. 1294 – 1301, Aug 2007.

[9] N. M. Pindorija, S. N. Singh, and S. K. Singh. "An adaptive wave neural network-based energy price forecasting in electricity markets," IEEE Trans. Power Syst., vol 23, no 3, pp. 1423 – 1432, Aug 2008.

[10] N. Amjady and H. Hemmati, "Day-ahead price forecasting of electricity markets by a hibrid intelligent system," Eur Elect. Power, vol 19, no 1, pp. 89 – 102, Jan 2009.

[11] M. Hanmandlu, B. K. Chauhan, "Load Forecasting Using Hybriid Models," IEEE Trans. Power Syst., vol 26, no 1, pp. 20 – 29, Feb 2011

[12] J. P. S. Catalao, H. M. I. Pousinho, V. M. F. Mendes, "Hybrid Wavelet-PSO-ANFIS Approach for Short-Term Electricity Price Forecasting," IEEE Trans. Power Syst., vol 26, no 1, pp. 137 – 144, Feb 2011

[13] A. Kale, 'Data preprocessing', hackernoon. Avaliable: https://hackernoon.com/data-preprocessing-85175b7a3f17

[14] Z. Weixiong, Data Mining. Avaliable: http://www.cs.wustl.edu/~zhang/teaching/cs514/Spring11/Data-prep.pdf

[15] XenonStack web-page, 'Data Preprocessing and Data Wrangling in Machine Learning and Deep Learning': https://medium.com/@xenonstack/data-preprocessing-and-data-wrangling-in-machine-learning-and-deep-learning-2122bc8daea9

[16] R. Petrichenko, K. Baltputnis, A. Sauhats, D. Sobolevsky, "District Heating Demand Short-Term Forecasting". 17th International Conference on Enviroment and Electrical Engineering (IEEE EEEIC 2017), Italy, Milan, 6-9 June 2017

[17] the European Centre of Medium-Range Weather Forecastst: https://www.ecmwf.int/

[18] P. Ivanova, A. Sauhats, O. Linkevics. Towards optimization of combined cycle power plants' start-ups and shut-down. Proceedings of 57th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON 2016), Latvija, Riga, 13.-14. October 2016

[19] R. Silipo, P. Winters, "Big Data, Smart Energy, and Predictive Analytics. Time Series Prediction of Smart Energy Data," report on data analytics based on the KNIME platform: https://files.knime.com/sites/default/files/inline-images/knime_bigdata_energy_timeseries_whitepape