District Heating Demand Short-Term Forecasting

Roman Petrichenko, Karlis Baltputnis, Antans Sauhats, Dimitry Sobolevsky
Faculty of Power and Electrical Engineering
Riga Technical University
Riga, Latvia
romans.petricenko@rtu.lv, karlis.baltputnis@rtu.lv. sauhats@eef.rtu.lv, dmitrijs.sobolevskis@rtu.lv

Abstract— **This paper discusses various forecasting tools that can be used in predicting the thermal load in district heating networks, focusing on day-ahead hourly planning as it is particularly important for cogeneration plants participating in electricity wholesale markets. Forecasts obtained by employing an artificial neural network are compared to a polynomial regression model. Their ability to supplement each other in a combined forecasting tool has been considered as well. Prediction inaccuracy cost is observed and suggested as evaluation criterion. The case studies are based on the district heating network in Riga, Latvia. Recorded data sets of temperature and heat demand are applied for thermal load prediction.**

Keywords— disctrict heating; forecasting; artificial neural networks; regression

I. INTRODUCTION

Energy market liberalization in various European countries brought fundamental changes to the structure and interactions between the different actors within their power systems, i.e., power producers, distribution and transmission system operators, wholesalers, retailers, and consumers. Increased competition is the main consequence of this process.

Those market participants that strive to maintain their profitability have to continuously invest in their technology, software solutions and planning approaches in order to not be sidelined by more effective competitors. One of the most important short-term planning tasks a generating company (GENCO) must solve when participating in a power exchange is obtaining adequately accurate consumption forecasts. This includes both, electricity consumption and heat demand, if the GENCO has cogeneration plants in their portfolio as their abilities to produce electrical power then are closely related to district heating demand in their supply area.

The current study is focused on the second part of that task – improving the accuracy of heat demand forecasts by comparing different methodologies and synthesizing one that offers better results. Authors present expenses calculation example of thermal load prediction inaccuracy.

There are several power systems that have a large influence of combined heat and power plants (CHPP), especially in Northern Europe [1], particularly – Latvia [2].

Various approaches of heat demand forecasting have been studied. For instance, in [3] a simple model is offered where heat consumption is divided in a component dependent on the ambient temperature and a second part that characterizes consumer behavior. The former is defined as a piecewise linear function obtained by solving a linear least squares problem, but the latter is derived from weekly patterns in heat consumption. The authors of [3] found that their results are of comparable accuracy to those obtained by more sophisticated commercial forecasting tools that relay on ARMA models. Similarly, in [4], the authors also used two components. The social component assumed a daily character and it was found by means of Box-Jenkins methodology.

A recent study based on the demand profile of Espoo in Finland found a linear regression model with a two-week social component to be the most useful in practice [5].

The authors of [6] compared ARIMA, EWMA (exponentially weighted moving average), linear regression and artificial neural network (ANN) models for thermal load forecasting. They concluded that of the models contrasted the ANN produced the most accurate results. A number of different studies have confirmed the applicability of ANN in heat demand forecasting at both, domestic level [7] and the whole district heating (DH) network [8], [9].

The authors of the current paper have previously analyzed ANN use in predicting electricity market prices and water inflow in reservoirs [10]. The goal now is to test an ANN based thermal load prediction model and compare it to one built on polynomial regression. The secondary objective is to evaluate the benefits that might arise if these models are combined. Predicted thermal load inaccuracy cost minimization is the final goal of this research. The case studies are based on the district heating network in Riga, Latvia, particularly, the zone supplied by cogeneration plants.

The paper is organized as follows. Section II provides an overview of the district heating (DH) system in Riga. In section III the ANN approach for thermal load forecasting is briefly described. Section IV represents thermal load forecasting based on a polynomic equation. Section V reports the results obtained by combining the methods and forecast inaccuracy cost estimation. Section VI reviews main results and Section VII concludes the article.

II. DISTRICT HEATING SYSTEM IN RIGA

A peculiarity of the DH network in Riga is the fact that it is effectively divided in several parts, furthermore, the heating demand in the supply area in the right bank of the city is covered by two large cogeneration plants – CHPP-1 and

CHPP-2 with electrical capacity 144 MWel, 832 MWel and thermal capacity 493 MWth, 544 MWth respectively. Fig. 1 shows annual heat delivered to consumers [11].

Aforementioned CHPPs mostly operate in the highly efficient cogeneration mode (generation of heat and power), while the largest plant also is capable of operating in mixed and condensing modes. The Riga CHPPs use natural gas (the least ecologically harmful of fossil fuels) as their primary fuel.

The schedule of electricity production in these cogeneration plants depends largely on thermal energy demand, which in turn depends on the climate conditions and duration of the heating season. The other important factor in CHP scheduling is the situation in electricity wholesale market, i.e., the day-ahead price signals.

Cabinet regulation No. 388 enforces Latvian construction standard LBN 003-15 "Construction climatology", which envisions the average (design) temperature in Riga to be 0.00 °C and the length of a heating season to be about 203 days. Fig. 2 and fig. 3 shows average seasonal ambient temperature and duration of the heating season from statistical data.

Climate conditions in Riga can vary strongly both, during the day and throughout the heating season. Fig. 4 illustrates the ambient temperature range in 24 hour periods (thermal load planning horizon for day-ahead, i.e. from 8.00 till 7.00 of next day – a peculiarity of the practices in the particular DH operational structure). As can be seen from fig. 4, the maximum daily ambient temperature range is 11.3 °C, it corresponds to January 7–8, 2017. The hourly minimum and maximum values within the aforementioned day are -18.2 °C and -6.9 °C, respectively [11].

III. HEAT DEMAND FORECASTING BY ANN

When setting up an ANN, several important decisions must be made. Firstly, it must be decided what time-series to use as input for training and forecasting. This largely depends on what processes are being predicted.

According to correlation analysis a crucial input variable for heat demand forecasting is the prediction of ambient temperature. Fig. 5 illustrates the result of the analysis: the correlation coefficient between load and temperature was 0.96 for the 2015/16 heating season.

In this study, we use the hourly temperature forecasts in Riga prepared by the Latvian Environment, Geology and Meteorology Centre. Training is performed on data sets of recorded temperature and demand. The type and structure of the ANN must also be selected. The authors employ narxnet – a nonlinear autoregressive neural network with external input [12] as it is implemented in the ANN toolbox of MATLAB 2013b.

The decision on the structure of an ANN is largely case dependent. For example, in [13] the authors concluded that, unfortunately, a generalized solution to this problem is not possible and the best network architecture for each particular application has to be determined experimentally.
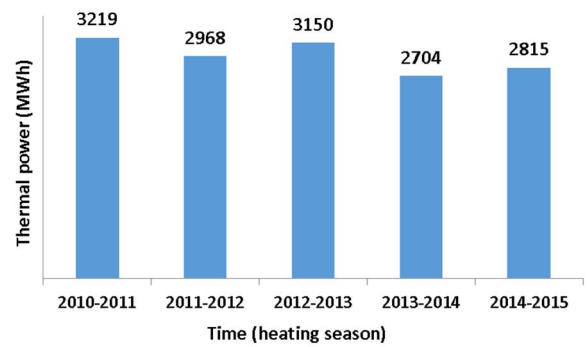

Fig. 1. Heat delivered to consumers.
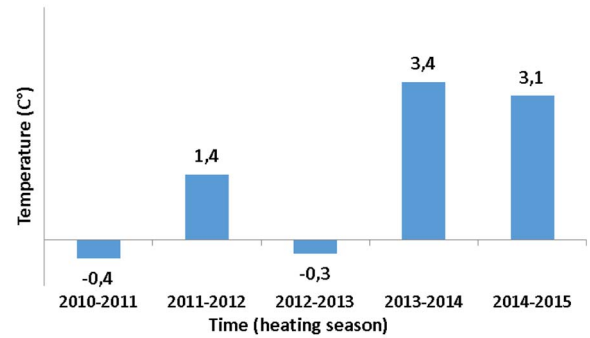

Fig. 2. Average ambient temperature during the heating season.
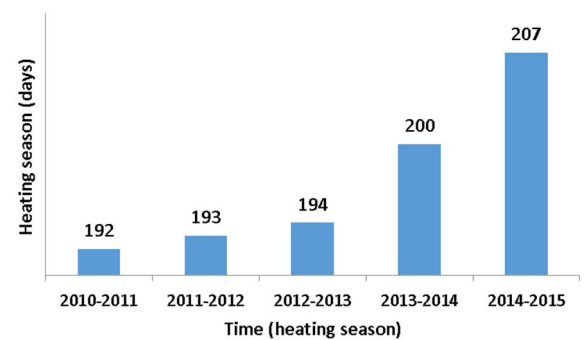

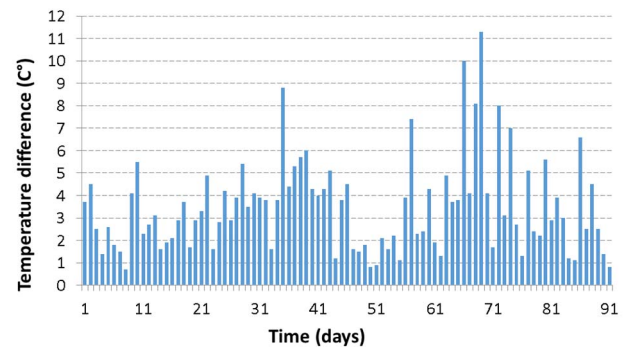Fig. 3. Duration of the heating season.

Fig. 4. Daily ambient temperature range from October 31, 2016 to January 31, 2017.
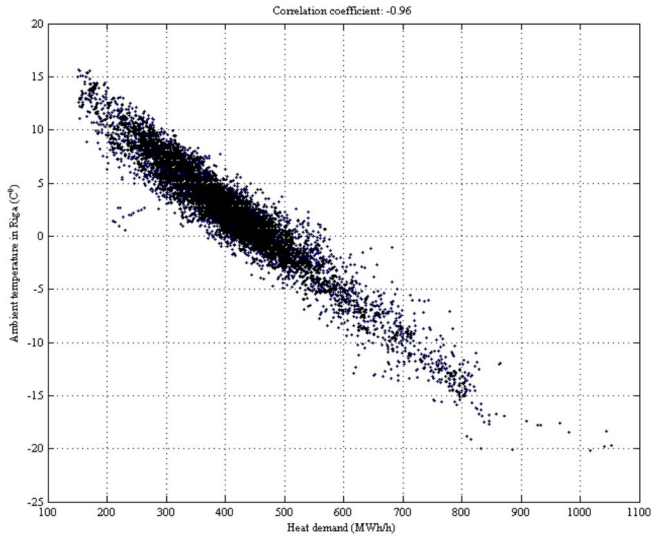
Fig. 5. Ambient temperature and heat demand correlation analysis

In [14] methods to improve the performance of ANNs are compiled. In the current study the authors apply some of them. For instance, one very valuable advice is the cyclical learning process. The result of such a process is a set of trained ANNs and their forecasts.

Finally, the structure of the network used in this study is as follows: input data set – 197 days, input delay – 1:4, feedback delay – 1, 50% of the data set is to be used for training, 10% for validation and the remaining 40% – for testing.

It should be noted that the separation of the data set into the various subsets is implemented randomly. As a consequence, the same input series may produce different forecasts. To alleviate this, we repeat the forecasting process 200 times and take the mean for each hour, since the forecast is performed for the day-ahead horizon (24 hours).

IV. Heat demand forecasting by polynomial regression

The second forecasting method studied in this paper utilizes the polyfit function of MATLAB 2013b (1) to find the polynomial coefficients (2). They are calculated by solving a least squares problem. Similarly to the previous method, the main input variable is ambient temperature.

$$p = polyfit(x, y, n), \quad (1)$$

$$p(x) = p_1 x^n + p_2 x^{n-1} + ... + p_n x + p_{n+1}, \quad (2)$$

where p – is a set of coefficients corresponding to input series x and output series y for a polynomial of the nth degree [15].

Additionally, a peculiarity of our approach is the fact that to account for a possible periodicity in the heat demand (e.g., caused by hot water supply or any other exogenous factors) we try to identify it by applying the hourly mean of the residuals for every of the 24 hours to the resulting forecast. For this method, the curve is fitted on the previous 8 days.

Generally, we used 3rd degree polynomials, as mostly they offered better results than other fits. However, high order polynomials carry a higher risk of unusable results if they have to extrapolate data. This lead to us adding a conditional operation in the software tool we developed – if the input forecasted temperatures for the next day were outside of the scope of the previous data the curve fitting function was switched to a first order polynomial.

V. Heat Demand Forecast by a Combined Method

As we expected both of the previously described methods to have their strengths and weaknesses, e.g., regression should clearly capture heat dependency on temperature, whereas an ANN might be better at identifying peculiarities in that characteristic. Besides, the ANN is set up in such a way as to consider inertia, i.e., lag in time between the input and output.

With that in mind we devised a third forecasting method, which produces results by taking the hourly mean between forecasts obtained by the first and second method. We hypothesize that this approach should provide better accuracy than each of the other methods independently.

All the three methods were implemented in MATLAB and tested on the historical heat demand and temperature data of December, 2016, in the DH network of Riga, right bank of the river Daugava. This is the supply zone provided with heat mainly by two cogeneration plants that sell their produced electricity in a market exchange (CHPP-1 and 2). Hence the importance of sufficiently accurate heating forecastsю

The forecasts are performed in a rolling manner, i.e., done for the next day-ahead horizon by updating the statistics with the actual load and temperature profiles of the previous day. The purpose of this is to closely simulate the actual use of forecasting software when it is deployed on site and used by cogeneration plant planning engineers.

VI. Results

Fig. 6 illustrates the accuracy of the results of all three forecast methods. The text boxes display the daily mean average percentage error (MAPE) (3) of the combined method.

Evidently, polynomial regression mostly offers better accuracy than ANN, however, there are days when the opposite is true. The combined method provides the smallest errors if we consider the whole one month horizon. In numerical terms, MAPE of the ANN forecasts was 5.83%, polynomial forecasts – 4.98%, and, finally, the combined forecasts – 4.76%.

$$MAPE = \frac{1}{m} \cdot \sum_{i=1}^{m} \left(\frac{|y_i - y_i'|}{y_i} \cdot 100 \right), \quad (3)$$

where m – total number of forecasted data points, y – historic data, y' - forecast, i – time step (in our case i equals one hour).

For comparison's sake, let us take a closer look at two particular days – one where polynomial regression provided better results (fig. 7), and one where the ANN did (fig. 8). It should be noted that the forecasts are performed for a daily

horizon from 8 am to 7 am to more accurately resemble the daily operations of planning engineers.

In the first example MAPE for the ANN forecast is 3.88%, polynomial – 3.63%, but combined method – 2.99%. Whereas in the second example MAPE for the ANN forecast is 3.8%, polynomial – 5.04% and combined method – 2.77%.

A. Financial impact of heating demand forecasting accuracy

For an example of the cost calculation of thermal load forecasting inaccuracy, let us consider the time series summarized in fig. 7. Fig. 9 displays the difference (Δ_{diff}) of the actual and predicted hourly values, this difference can be both, positive and negative.

If the CHPPs operate solely in cogeneration mode and do not have significant thermal energy storage options, as is currently the case in Riga, their electrical output is proportional to thermal energy production. Thereby, if Δ_{diff} is negative there is a deficit of thermal power which is alleviated by increasing production. Consequently, electricity production increases as well, but the new volume had not been sold in the day-ahead market. If Δ_{diff} is positive the actual heating demand is less than planned and the CHPP does not produce as much electricity as it had sold, unless it opts to waste the heating overproduction.

Abovementioned two cases lead to financial loses. Let us assume coefficients to estimate the costs associated with under- and overproduction of energy, $\beta = 2$ and $\beta = 1/3$ respectively.

$$\begin{cases} E = \sum_{i=1}^{n} \left| \Delta Q_i \cdot \alpha \right| \cdot \beta_i \cdot C_i \\ \Delta_{dif_i} < 0 \rightarrow \beta_i = 2 \\ \Delta_{dif_i} = 0 \rightarrow \beta_i = 1 \\ \Delta_{dif_i} > 0 \rightarrow \beta_i = \frac{1}{3} \end{cases}, \qquad (4)$$

where E – expenses, ΔQ_i – thermal load value at i-th hour, α – constant coefficient to express produced electrical energy from produced heat, 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.

System of equation (4) should be explained practically with the previously mentioned data from fig. 7. Table I lists data for some particular hours, identified in column 'T'. The historical heating demand is summarized in column 'ATT' and forecast obtained by the 1st method (ANN) in the column labeled 'ANN'. The 'Diff' column displays forecast errors. 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) [16]. Column labeled 'D' shows the status of the predicted thermal load in regard to the actual.

The corrected electricity prices (column 'CP' in €/MWh) depend from equation (4) and the direction of the forecast error.
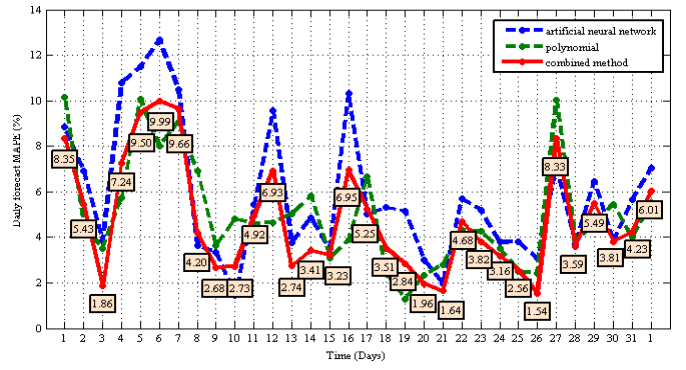

Fig. 6. Thermal load prediction accuracy for the first, second and combined methods
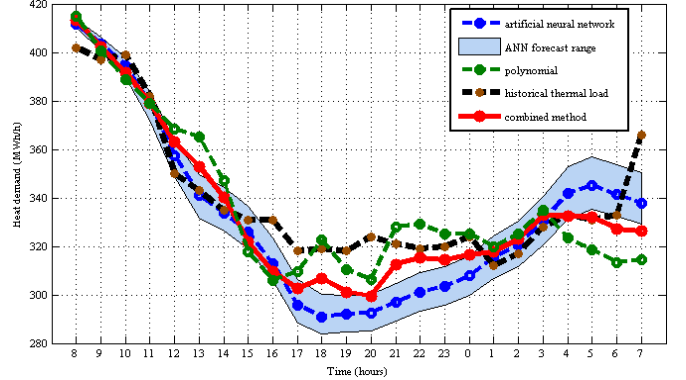

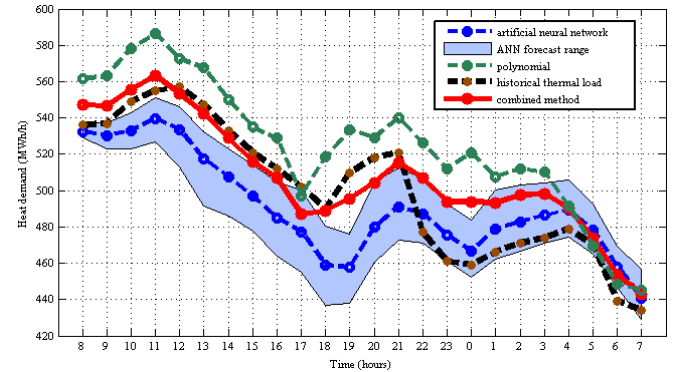Fig. 7. Three methods of thermal load forecasting, 8–9 December, 2016

Fig. 8. Three methods of thermal load forecasting,12–13 December, 2016

It is thus possible to calculate the expenses caused by forecast inaccuracies (column 'E' in €). The bottom-right cell in Table I shows the sum of hourly estimated expenses incurred due to imperfect forecast. For the selected day (December 8–9), the total expense estimate reaches 29 102.35 €. 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 by the polynomial and combined methods.

TABLE I. INACURACY COST OF HEAT DEMAND FORECAST BY ANN

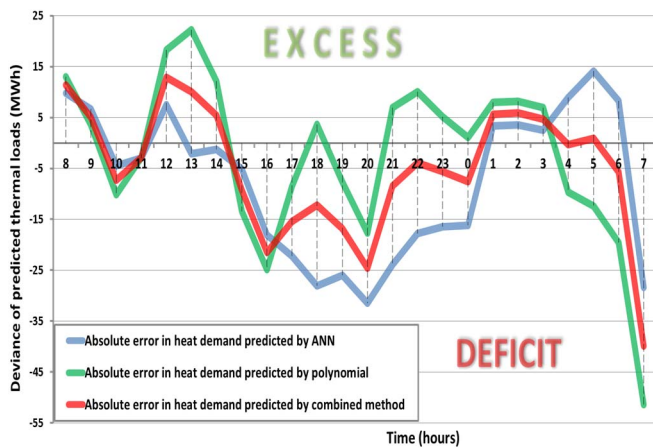| T	ATL	ANN	Dif	EP	D	CP	E
8	402	412	-14,70	46,09	>	15,36	225,89
9	397	404	-9,95	41,07	>	13,69	136,18
…	…	…	…	…	…	…	…
6	333	341	-12,36	40,76	>	13,59	167,89
7	366	338	42,49	45,97	<	91,94	3906,16
							Sum
							29 102.35 (€)

Fig. 9. Inaccuracies in heat demand forecast for 8-9 December 2016

Fig. 9 displays forecast errors obtained by ANN, polynomial and combined methods. Based on these results it is possible to calculate the inaccuracy cost of thermal load forecasting by the considered methods. Table II summarizes the results.

Combined method shows the lowest forecast inaccuracy – 2.99%. It also provides the lowest value of the thermal load prediction inaccuracy cost – 23 164.68 €.

TABLE II. FORECAST ERRORS AND THE CORRESPONDING COSTS FOR THE THREE STUDIED METHODS

Date	ANN	Polynomial	Combined
Dec. 8–9, 2016	MAPE (%)	MAPE (%)	MAPE (%)
	3.88	3.63	2.99
	PIC (€)	PIC (€)	PIC(€)
	29 102.35	24 309.35	23 164.68

PIC – prediction inaccuracy cost.

We can easily conclude that thermal load prediction errors are expensive. This signifies the value to try and increase the accuracy of forecasting tools.

VII. CONCLUSIONS

A district heating network thermal load demand forecasting approach based on polynomial regression proved to be overall more accurate than an ANN based method for the case study of Riga right bank DH network. However, if the curve fitting approach is supplemented with the ANN forecasts, the mean results provided a small but noticeable improvement in the error measures, pointing to benefits in the combined application of both methods.

The expenses incurred by the operator of DH supplying CHPP operators due to heating demand forecast inaccuracies was roughly estimated as well. A conclusion can be drawn that these errors can indeed be costly and thus a further search for ways to improve forecasting accuracy is justified.

REFERENCES

[1] Combined heat and power (CHP). [Online]. Available: http://www.eea.europa.eu/data-and-maps/indicators/combined-heat-and-power-chp-1/combined-heat-and-power-chp-2. [Accessed: 03-Apr-2017].

[2] International Energy Agency (EIA), "Cogeneration and District Energy," p. 60, 2009.

[3] E. Dotzauer, "Simple model for prediction of loads in district-heating systems," *Appl. Energy*, vol. 73, no. 3–4, pp. 277–284, Nov. 2002.

[4] B. Chramcov, "Heat demand forecasting for concrete district heating system," *Int. J. Math. Model. Methods Appl. Sci.*, vol. 4, no. 4, pp. 231–239, 2010.

[5] T. Fang, R. Lahdelma, "Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system," *Appl. Energy*, vol. 179, pp. 544–552, Oct. 2016.

[6] M. Kawashima, C. E. Dorgan, and J. W. Mitchell, "Hourly thermal load prediction for the next 24 hours by Arima, Ewma, LR, and an artificial neural network," American Society of Heating, Refrigerating and Air-Conditioning Engineers, 1995.

[7] V. Bakker, A. Molderink, J. L. Hurink, and G. J. M. Smit, "Domestic heat demand prediction using neural networks," *Systems Engineering*, pp. 189–194., 2008.

[8] W. Schellong and F. Hentges, "Forecast of the heat demand of a district heating system," *Proceedings of European Power and Energy Systems*, pp. 383–388, 2007.

[9] K. Wojdyga, "Predicting Heat Demand for a District Heating Systems," *Int. J. Energy Power Eng.*, vol. 3, no. 5, p. 237, 2014.

[10] A. Sauhats, R. Petrichenko, Z. Broka, K. Baltputnis, and D. Sobolevskis, "ANN-Based Forecasting of Hydropower Reservoir Inflow," Power and Electrical Engineering of Riga Technical University (RTUCON), 2016 57th International Scientific Conference on, pp. 2–7, 2017.

[11] JSC Latvenergo combined heat and power plants. [Online]. Available: http://www.latvenergo.lv/eng/about_us/generation/chpps/. [Accessed: 03-Apr-2017].

[12] "MATLAB." The MathWorks Inc, Natick, Massachusetts.

[13] T. Kavzoglu, "Determining Optimum Structure for Artificial Neural Networks," *proceddings 25th Annu. Tech. Conf. Exhib. Remote Sens. Soc.*, no. September, pp. 675–682, 1999.

[14] Improve neural network generalization and avoid overfitting, 2016. [Online]. Available: https://www.mathworks.com/help/nnet/ug/improve-neural-network-generalization-and-avoid-overfitting.html. [Accessed: 01-Feb-2017].

[15] Polyfit, 2016. [Online]. Available: https://www.mathworks.com/help/matlab/ref/polyfit.html. [Accessed: 01-Feb-2017]

[16] Nord Pool electrical energy price statistics. [Online]. Available: http://www.nordpoolspot.com/Market-data1/Elspot/Area-Prices/LV/Hourly/?view=table. [Accessed: 03-Apr-2017].