Comparative Study of Week-ahead Forecasting of Daily Gas Consumption in Buildings Using Regression ARMA/SARMA and Genetic-algorithm-optimized Regression Wavelet Neural Network Models

Alexander Hošovský, Ján Piteľ, Milan Adámek, Jana Mižáková, Kamil Židek

PII: S2352-7102(20)33587-7

DOI: https://doi.org/10.1016/j.jobe.2020.101955

Reference: JOBE 101955

To appear in: *Journal of Building Engineering*

Received Date: 13 April 2020

Revised Date: 19 September 2020

Accepted Date: 29 October 2020

Please cite this article as: A. Hošovský, J. Piteľ, M. Adámek, J. Mižáková, K. Židek, Comparative Study of Week-ahead Forecasting of Daily Gas Consumption in Buildings Using Regression ARMA/SARMA and Genetic-algorithm-optimized Regression Wavelet Neural Network Models, *Journal of Building Engineering*, https://doi.org/10.1016/j.jobe.2020.101955.

Comparative Study of Week-ahead Forecasting of Daily Gas Consumption in Buildings Using Regression ARMA/SARMA and Genetic-algorithm-optimized Regression Wavelet Neural Network Models

Alexander Hošovský[a*], Ján Piteľ[a], Milan Adámek[b], Jana Mižáková[a], Kamil Židek[a]

[a] *Faculty of Manufacturing Technologies with seat in Prešov, Technical University of Košice, Bayerova 1, 08001 Prešov, Slovakia*

[b] *Faculty of Applied Informatics, Tomas Bata University in Zlín, Nad Stráněmi 4511, 76005 Zlín, Czech Republic*

** - corresponding author*

Abstract

Forecasting energy consumption in buildings is crucial for achieving effective energy management as well as reducing environmental impacts. With the availability of large amounts of relevant data through smart metering, gas consumption forecasting is becoming an integral part of smart building design so that these requirements are met. In this study, we investigate week-ahead forecasting of daily gas consumption in three types of buildings characterized by different gas consumption profiles during a five-year period. As gas consumption in buildings is highly correlated with the average outdoor temperature, regression models with additional residual modeling are used for forecasting. However, conventional regression models with autoregressive moving averages (ARMA) errors (regARMA) perform poorly when the temperature forecasts are inaccurate. To address this, a new forecasting model termed genetic-algorithm-optimized regression wavelet neural network (GA-optimized regWANN) is proposed. It uses the wavelet decomposition of the residuals of temperature regression time-series, which are modeled by multiple nonlinear autoregressive (NAR) models based on sigmoid neural networks. The appropriate delays in the regression vectors of the NAR models are selected using a binary GA. Compared with regARMA and seasonal regARMA, the GA-optimized regWANN model achieved in the three buildings a reduction of 22.6%, 17.7%, and 57% in the mean absolute error (MAE) values in *ex post* forecasting with recorded temperatures, and a 52.5%, 27%, and 43.6% reduction in the MAE values in *ex ante* forecasting with week-ahead forecasted temperatures, even under conditions of relatively significant errors in the forecasted temperature.

Keywords: wavelet transform, neural networks, ARMA models, forecasting accuracy, temperature regression

Comparative Study of Week-ahead Forecasting of Daily Gas Consumption in Buildings Using Regression ARMA/SARMA and Genetic-algorithm-optimized Regression Wavelet Neural Network Models

Alexander Hošovský[a*], Ján Piteľ[a], Milan Adámek[b], Jana Mižáková[a], Kamil Židek[a]

[a] *Faculty of Manufacturing Technologies with seat in Prešov, Technical University of Košice, Bayerova 1, 08001 Prešov, Slovakia*

[b] *Faculty of Applied Informatics, Tomas Bata University in Zlín, Nad Stráněmi 4511, 76005 Zlín, Czech Republic*

** - corresponding author*

Abstract

Forecasting energy consumption in buildings is crucial for achieving effective energy management as well as reducing environmental impacts. With the availability of large amounts of relevant data through smart metering, gas consumption forecasting is becoming an integral part of smart building design so that these requirements are met. In this study, we investigate week-ahead forecasting of daily gas consumption in three types of buildings characterized by different gas consumption profiles during a five-year period. As gas consumption in buildings is highly correlated with the average outdoor temperature, regression models with additional residual modeling are used for forecasting. However, conventional regression models with autoregressive moving averages (ARMA) errors (regARMA) perform poorly when the temperature forecasts are inaccurate. To address this, a new forecasting model termed genetic-algorithm-optimized regression wavelet neural network (GA-optimized regWANN) is proposed. It uses the wavelet decomposition of the residuals of temperature regression time-series, which are modeled by multiple nonlinear autoregressive (NAR) models based on sigmoid neural networks. The appropriate delays in the regression vectors of the NAR models are selected using a binary GA. Compared with regARMA and seasonal regARMA, the GA-optimized regWANN model achieved in the three buildings a reduction of 22.6%, 17.7%, and 57% in the mean absolute error (MAE) values in *ex post* forecasting with recorded temperatures, and a 52.5%, 27%, and 43.6% reduction in the MAE values in *ex ante* forecasting with week-ahead forecasted temperatures, even under conditions of relatively significant errors in the forecasted temperature.

Keywords: wavelet transform, neural networks, ARMA models, forecasting accuracy, temperature regression

1 Introduction

In general, the prediction of energy use is currently highly relevant with regard to improving the energy performance in buildings as well as reducing the potential environmental impact. Even though short forecasting horizons may be relevant in certain applications areas (e.g., electrical energy), the total daily energy consumption (including, for example, gas) can also be important for effective energy management. The conditions of a particular application usually determine the suitability of a forecasting model. Various approaches have been successfully used for energy-use prediction in different types of buildings [1]. Although a universally accepted classification may not be available, three approaches are distinguished in this study: prediction using engineering, statistical, and artificial intelligence (AI) methods. As pointed out in [1], engineering methods use standard calculations based on known physical laws (thermodynamics and energy behavior), thus necessitating a large number of physical variables and parameters. In contrast, the other approaches are data-based, and the identified prediction models have no relation to the underlying physical laws (except for the possible use of, for example, meteorological variables as predictors). A similar but slightly more refined classification can be found in [2], where data-driven and large-scale building-energy-based approaches (LSBE) are distinguished. The former are further classified into approaches based on neural networks (NNs), clustering, statistical methods and machine learning, and support vector machines (SVMs), whereas the latter into black-, gray-, and white-box model approaches. A more focused review of energy models can be found in [3], where the emphasis is on energy demand forecasting specifically through data-based approaches that fall under the aforementioned statistical and AI categories. It should be noted that despite the distinctive characteristics of time series for different types of energy (e.g., thermal, electrical, or wind), the general characteristics of the methods used for time-series forecasting in these areas are similar and can be used as a reference from one field to another. The authors of [4], compared three machine learning approaches (feedforward NNs, regression trees, and SVMs) to forecast the thermal load in several residential and nonresidential buildings. The results demonstrated that the SVM approach was superior to the other two in terms of the normalized root mean square (NRMSE) value. Neural networks were also used in [5]. Specifically, they were trained to forecast daily energy consumption in buildings based on data classification and a direct multistep approach. A simple type of feedforward NN was used to forecast the energy load in an iterative manner based on the load during the previous five days. However, [5] is specifically concerned with energy forecasts in institutional buildings, which have distinctive energy profiles compared with other types of buildings. Taspinar et al. [6] used NNs for short-term forecasting of natural gas consumption; several meteorological variables (moisture, atmospheric pressure, wind speed, and

ambient temperature) were added to improve accuracy. The analyzed time series exhibited a strong seasonality of seven lags, and thus the SARIMAX (seasonal autoregressive moving average) model with exogenous input was used as a reference. It was demonstrated that the performance of SARIMAX was highly satisfactory and, in fact, superior to that of NN models. In [7], an NN model based on a multiobjective genetic algorithm (GA) was used for energy consumption prediction in a bioclimatic building. Here the authors used MOGA to design RBF models for 4-steps ahead forecasting the electric power consumption in selected buildings. Their model used much lower number of samples then the models to which authors compared the performance (2592 vs. 318340). However, even this number of samples may be high for daily samples-based forecasting of energy consumption. Jovanovic et al. [8] proposed an ensemble of three different architectures (FFNN, RBF, and ANFIS) for the prediction of heating energy consumption. The outputs of these models were combined by simple and weighted averaging, and by calculating the median. Several meteorological variables (outside mean daily temperature, heating consumption of the previous day, day of the week, maximum daily temperature, relative humidity, total solar radiation, and month of the year) were considered. However, it appears that only a short forecasting horizon was used, and the effect of seasonality was ignored.

Regardless of model architecture, forecasting accuracy may be improved by using advanced signal processing techniques, which may either extract useful features from the predicted time series or perform decompositions to decrease forecasting errors. A signal processing method is the wavelet transform (WT), which is widely used in time series analysis because of its locality not only in the frequency domain but also in the time domain [9]. Rana and Koprinska [10] used this approach in the form of a wavelet NN, where the entropy cost function was used to select the best wavelet basis for data decomposition, a mutual information criterion for feature selection, and the NN itself for forecasting. The combination of wavelet decomposition with autoregressive (integrated) moving averages (AR(I)MA) models was proposed in [11]. This led to an improvement in forecasting accuracy for short-term forecasts of PM10 concentration. Further improvements were achieved in [12] by using wavelet decomposition and an NN trained by a criss-cross optimization algorithm, which reportedly had a significant advantage over typical training algorithms. A WT-based approach was also used in [13], where it was combined with three other computational intelligence paradigms (ANFIS, NNs, and GA) to perform one-day-ahead natural-gas demand forecasting. In [13], a large number of towns were analyzed, but a comparison with well-established forecasting models was missing; furthermore, weather conditions were not considered. The ability of wavelets to detect non-stationary features of time series was combined with nonlinear Volterra models to provide a forecasting framework for nonlinear and non-stationary time series in [14]. This combination was reported to

achieve better results than certain benchmark models; however, Volterra models are high complex, and this should be considered.

Recently, time-series forecasting has also been addressed using deep learning approaches [15, 16, 17, 18, 19]. In [15], multi-layer Bi-LSTM and LSTM with a GA were combined for hourly forecasts of natural-gas demand, where the GA was used for network optimization. No meteorological variables were included in the model, but differences between summer and winter seasons were considered. The number of samples used to train the proposed model were not reported, but it is conceivable that a large number is required for satisfactory performance. Liu et al. [16] used deep learning based techniques specifically for forecasting the energy consumption in buildings. In this work authors used less conventional approach for forecasting, where a number of deep reinforcement techniques was explored for the performance enhancement in single- and multi-step ahead prediction of the energy consumption. It was found out that one of the investigated techniques was not well-suited to this problem (Asynchronous Advantage Actor-Critic) but Deep Deterministic Policy Gradient and Recurrent Deterministic Policy Gradient performed better. This was, however, achieved at the cost of higher computational cost typically associated with the deep learning methods. To improve generalizability, a double deep extreme learning ensemble system with a self-adaptive ReTSP-Trend pruning technique was proposed in [17] and was demonstrated to perform better than state-of-the-art algorithms. Even though NNs with deep architecture appear to be highly effective for time-series forecasting, their effectiveness depends heavily on the use of large datasets, typically containing thousands of samples. Chitalia et al. [18] tested nine different combinations of recurrent NNs and clustering techniques for the short-term forecasting of electrical load in different types of commercial buildings. It was demonstrated that the combination of long short-term networks and clustering techniques can significantly improve forecasting accuracy compared with other techniques with good robustness towards errors in weather forecasts. However, the proposed models are complex and computationally intensive. Moreover, the time resolution is 1 h or 15 min, which is sufficient to generate large amounts of data in a relatively short time. Therefore, the method would not be suitable for forecasts with daily resolution.

Despite the large number of studies related to the use of AI methods in time-series forecasting, AR(I)MA models, either in pure form or hybridized with other models, are widely used for numerous time series encountered in practice [20]–[30]. Vaghefi et al. [20] used a generalized form of the Cochrane–Orcutt estimation technique, in which a multiple linear regression model was combined with a seasonal ARMA (SARMA) model to forecast cooling and electricity load demand in a cooling, heating, and power plant. Seasonal effects were also

considered by Akpinar and Yumusak [21], who compared the forecasting performance of two approaches (Holt–Winters exponential smoothing and ARIMA) for the year-ahead natural gas demand in Sakarya province (Turkey). It was demonstrated that the ARIMA model yielded the best results. An interesting approach was taken in [22] and [23], where short-term load forecasting was based on the similarity of patterns of seasonal cycles using various methods (kernel estimation, nearest neighbor estimation, and clustering). A comparison was made with other, more conventional forecasting methods (e.g., ARIMA, exponential smoothing, and NN), and the results demonstrated that the proposed method performed highly satisfactorily. In [24], the significant complexity of factors related to the energy behavior in buildings was addressed using a feature selection method based on support vector regression with two kernel functions (RBF and polynomial). It was demonstrated that this method has high prediction accuracy and relatively short computational time. In addition to wavelet analysis, Fourier analysis is also a powerful method with high potential for forecasting. Yukseltan et al. [25] used Fourier analysis with feedback for hourly electricity demand forecasting. The method was also combined with a conventional autoregressive model to slightly improve performance. The time series in that study exhibited strong seasonality, but the performance of the method on data with more complex seasonality was not clearly evaluated. Owing to the strong correlation of energy consumption with weather conditions, certain meteorological variables are occasionally used to achieve lower prediction errors. In addition to this, calendar effects can also be a factor that can be taken into account in forecasting energy load in residential buildings. Lusis et al. [26] analyzed the effects of calendar effects, forecasting granularity and training sets sizes on the forecasting accuracy. However, it was found that coupled with the weekly seasonality and weather data this specific factor may not have significant effect on the forecasting performance. As a result, more attention may be paid to the measurement of additional meteorological variables (ambient temperature, global radiation, and wind speed). These additional meteorological variables (ambient temperature, global radiation, and wind speed) were also considered in [27]. This was combined with adaptive linear time-series models, resulting in satisfactory forecasting performance of the load for a group of family houses in Denmark. Nevertheless, in many cases the range of monitored meteorological variables is limited and the outdoor temperature remains the most easily available variable of this type.

Several observations regarding energy forecasting can be made based on this literature review. Generally, electrical energy forecasting appears to be dominant in terms of the number of studies. Even though the methods used in the present study are similar, gas consumption in buildings has certain specific features that differentiate forecasting. Various techniques are used for forecasting, but statistical and AI-based methods appear to be

particularly attractive. It is important to note that ARMA/ARIMA models, which are the most commonly used forecasting models, can be advantageously applied to several real-world processes. Therefore, these models should serve as a reference for comparison with any AI method. Moreover, the recent trend of using deep learning methods in forecasting has brought several benefits, which, however, depend on the availability of large amounts of data. When short time intervals are of interest (e.g., hourly data), a sufficient number of samples can be obtained in a reasonable time. However, in the case of longer intervals (e.g., daily data), the parsimony of deep learning models may be compromised. It can also be observed that the use of wavelet decomposition in forecasting is relatively popular. However, the analysis carried out in [31] using the fast Fourier transform for the wavelet decomposition components of the gas consumption profiles in various buildings indicated the presence of complex seasonality, which should be considered in the development of forecasting models.

With regard to the purpose of forecasting week-ahead daily gas consumption in different types of buildings, the objectives of our work can be summarized as follows:

1. To develop a reference forecasting model using regARMA/SARMA in rigorous way so that possible improvements are compared to the best obtainable performance of a standard model
2. To propose the model with improved performance, the use of which will be simple and possible for heat sources where the possibilities for measuring various meteorological data are limited.
3. To present the systematic methodology for the identification of this model as well as forecasting of the week-ahead consumption in a possible online scenario.

To meet the objectives above and also address the aforementioned points identified in the knowledge gap a new forecasting model for the week-ahead gas consumption in different types of buildings is introduced. In addition to the temperature regression part, this model uses the Dabeuchies wavelet decomposition of regression model residuals, neural network-based NAR models for modeling the wavelet decomposition components and binary genetic algorithm for searching for near-optimal lags in NAR models regressor vectors. This arguably improves the performance of the (S)ARMA model and is relatively simple to implement given the availability of relevant data. Even though the inclusion of various meteorological variables may further improve forecasting accuracy [8, 26], we used only the average daily outdoor temperature in the proposed model. Therefore, in a real-world scenario, only the measurements of gas consumption and temperature, which can be easily obtained, are required. The model is intended to be simple to use once the identification has been carried out and to help with that a systematic methodology of its identification and use is presented. The data acquisition methods and

devices as well as the related data processing were partly based on previous experiments in biomass combustion monitoring and control [32], [33], whereas a general description of the approaches used in the monitoring of heat production and distribution under conditions similar to those in the present study was given in [34]. To address the issue of multiple seasonal cycles in the gas-consumption time series obtained from the monitored buildings [31], the wavelet decomposition approach was applied, and these cycles were included in the various detail components of the decomposed time series. Nonlinear autoregressive (NAR) models were used to approximate the WT components of the underlying process, as these models are easier to train (than (nonlinear) ARMA models); however, they usually have more parameters [35]. To determine suitable lags in the regressor vectors of the NAR models (this is related to the aforementioned complex seasonality of the gas consumption profiles of the examined buildings), a binary GA was applied. This enabled a near-optimal selection of the lags in the regressor vector for forecasting the temperature regression residuals.

The remainder of the paper is organized as follows. Section 2 contains basic information regarding the time series describing the total daily gas consumption profiles in each type of building, and the related analysis from an identification and/or forecasting perspective. Section 3 describes the relevant methods used for the identification of linear regression with (S)ARMA errors (reg(S)ARMA) and the GA-optimized regWANN. Section 4 contains the description and discussion of weak-ahead forecasting results obtained using both models for each building. Section 4.4 describes the most important observations based on the results and their significance. Section 5 presents a broader view of the results by drawing general conclusions and mentioning several aspects important for future work.

2 Data description

In this study, we are concerned with both residential and non-residential buildings: a healthcare building, an elementary school, and military quarters (also used as a hotel). They are located in the eastern/north-eastern part of Slovakia (Fig. 1). The basic data used in the experiments consisted of six separate datasets in total, containing daily records of the gas consumption (taken at 6.00 am for the previous day) and average outdoor temperature measured in the vicinity of the buildings. The total size of these datasets differed because the recording of the data became possible in a different period for each building. The raw time series corresponding to the recorded gas consumption in the given buildings is shown in Fig. 2, with the Tatranské Zruby dataset starting on December 22, 2013, and ending on December 15, 2018 (1820 consecutive days or samples), the Krompachy dataset starting on November 29, 2013, and ending on December 24, 2018 (1851 consecutive days or samples), and the Sečovce dataset starting on September 20, 2014, and ending on January 30, 2019 (1594 consecutive days

or samples). It was assumed that the gas consumption forecasting was useful only during the heating season (from October 1 to May 1; indicated by HT in the figure). Accordingly, the corresponding data were extracted from each time series.

Fig. 1. Buildings for week-ahead forecasting of daily gas consumption (left: hotel in Tatranské Zruby, middle: an elementary school in Krompachy, right: healthcare building in Sečovce, source: Google Maps).

By considering a leap day (February 29, 2015), three separate datasets with 213 samples (years 2014, 2016, and 2018), and one with 214 samples (year 2015) were obtained for each building. An additional dataset (TD in the figure) was formed in each case by extracting the data starting from October 1, 2018, until the end of the given (main) dataset to obtain test data not used in the identification process (76 samples for the Tatranské Zruby dataset, 86 samples for the Krompachy dataset, and 122 samples for the Sečovce dataset).

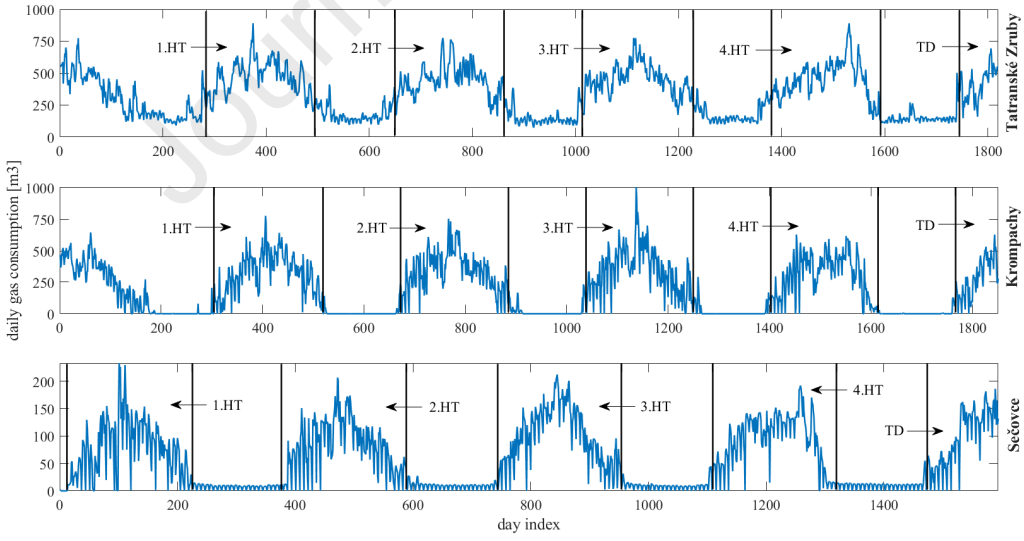


Fig. 2. Raw time series characterizing daily gas consumption in each building (HT: heating seasons in four years, TD: test data).

The extracted time series for the considered periods and the corresponding daily average outdoor temperatures are shown in Fig. 4. Moreover, basic data statistics regarding the time series can be examined using the box plots in Fig. 3. The actual gas consumption profiles for each building were affected by various factors, such as main

purpose and size, geographical location (manifested through different statistics of the outdoor temperature), and specific boiler operation conditions. The daily gas consumption in Sečovce was, on average, considerably lower (median: 88.57 m^3, maximum value: 233 m^3) than that in Krompachy (median: 336 m^3, maximum value: 1003.63 m^3) and Tatranské Zruby (median: 459.5 m^3, maximum value: 891 m^3). It is interesting to observe that the median values of the yearly gas consumption are almost equal in Krompachy; this is possibly because this building has a more stable operation (elementary school) than the other two. It should also be noted that, after the extraction of the relevant data, no samples were removed from any time series for identification and/or testing because none of the data points indicated as outliers (and calculated as $q_2 - 1.57(q_3 - q_1)/\sqrt{N}$, where q_1, q_2, q_3 are the 25th, 50th, and 75th percentiles, respectively, and N is the number of samples) were attributed to a failure in the measurement process, resulting in clearly incorrect gas consumption values. The largest number of statistical outliers was observed in Tatranské Zruby (27 points in total compared with 2 and 0 points for Krompachy and Sečovce, respectively). This is most probably related to the sudden changes in outdoor temperature (which might have been more severe in this region), but their distribution and number are not simply correlated with the outliers in the temperature data. The daily average outdoor temperatures are statistically in correspondence with the geographical location of the buildings, with Tatranské Zruby being the coldest location (median value 2.69°C, and max/min 17.47°C/−18.36°C), followed by Krompachy (median value 3.38°C, and max/min 18.50°C/−15.18°C) and Sečovce (median value 4.83°C, and max/min 23.28°C/−15.09°C).

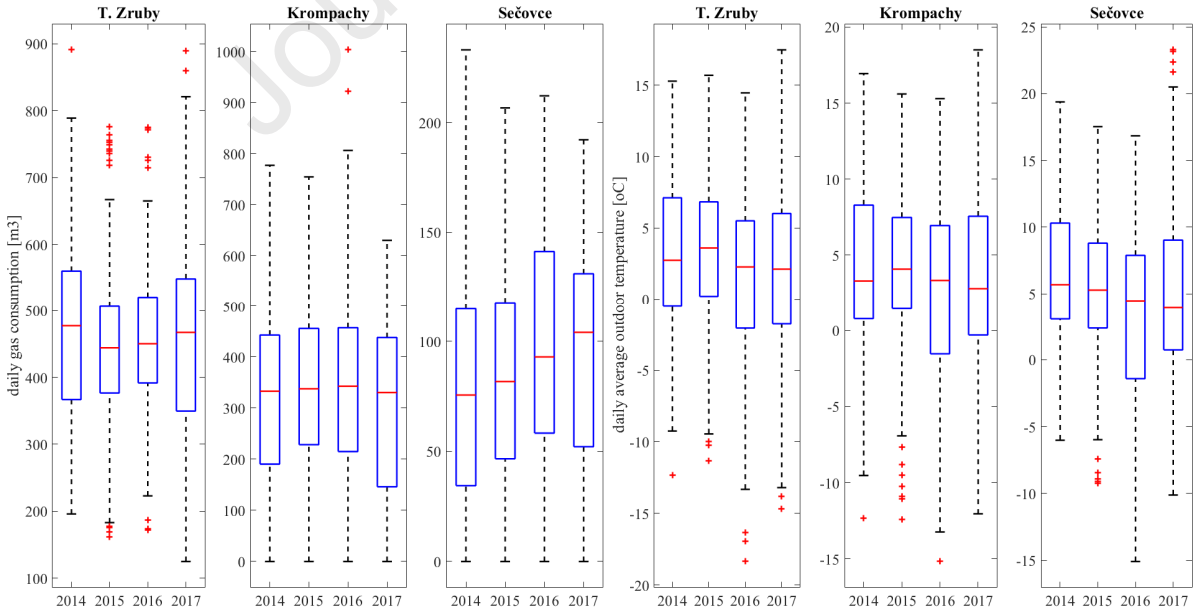


Fig. 3. Boxplots of daily gas consumption in given buildings and average daily outdoor temperatures at those locations for the heating season in years 2014, 2015, 2016, and 2017.

Visual inspection of the time series graphs in Fig. 4 indicates not only a significant correlation of the daily gas consumption with the outdoor temperature but also a strong seasonality. This is considerably more pronounced in the Krompachy and Sečovce datasets than in the Tatranské Zruby dataset, as the gas consumption in these buildings is characterized by typical weekly seasonal periods, with weekends being almost visitor-free (elementary school and healthcare building). The sudden drop in gas consumption close to the middle of the heating season is due to the holidays at that time. This has more pronounced in Krompachy (elementary school), as the other buildings continue to operate during these periods.

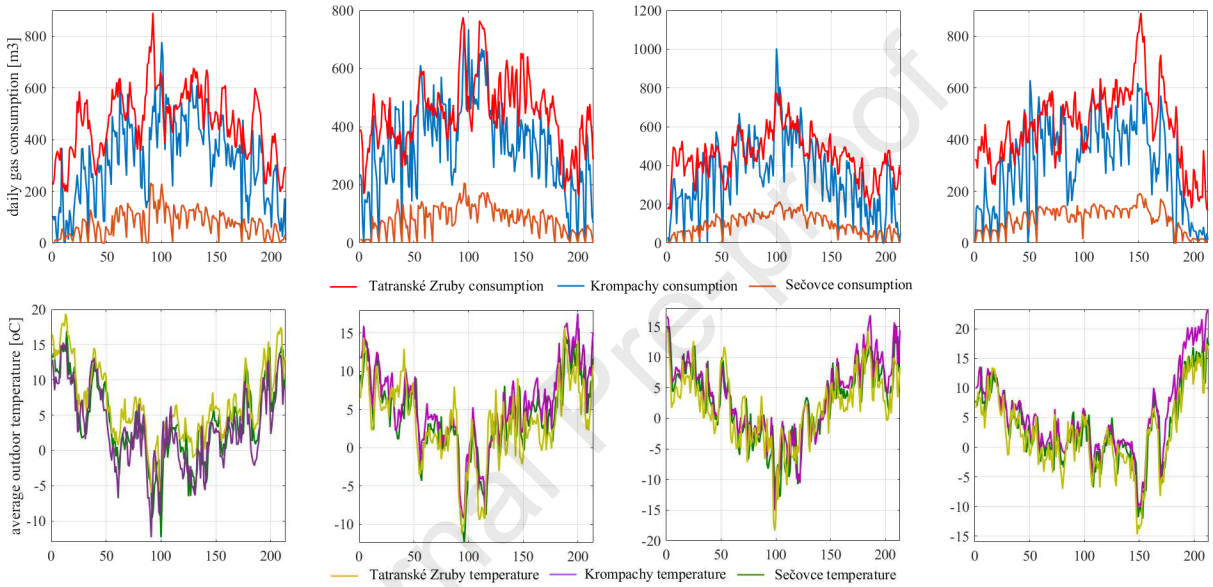


Fig. 4. Time series of daily gas consumption and average daily outdoor temperatures in given heating seasons.

3 Methods

Two types of models are used to forecast the week-ahead gas consumption in the three buildings. As the daily gas consumption in any these buildings is strongly correlated with the outdoor temperature, regression models are selected, where time-independent regression with temperature is used, and the time-series modeling approach is adopted for the residuals. The conventional SARMA model is used a reference with which the performance of the proposed model is compared. To improve the forecasting accuracy for week-ahead gas consumption, a GA-optimized wavelet NN forecasting model is introduced. This model uses a wavelet decomposition of the time-series residuals into approximation and detail components, each of which is modeled using an NAR model with a sigmoid NN. Lag selection in the regressor vectors of the NAR models is performed by using a binary GA so that a suitable combination may be obtained. The combination of temperature regression and NAR models with appropriate lag selection is then used to forecast the weak-ahead gas consumption in the selected buildings. In

the following section, reg(S)ARMA, the WT, and the binary GA are briefly reviewed. Subsequently, the methodology for identifying and using the GA-optimized regWANN forecasting model is introduced.

3.1 reg(S)ARMA forecasting model

It is readily observable from Fig. 4 that the daily gas consumption in all buildings is strongly correlated with the average daily outdoor temperature, suggesting the suitability of regression forecasting models. When the residuals are serially correlated after a regression model is applied, it is appropriate to use a regARMA approach, that is, a regression model with ARMA-modeled time-series error terms [35]. Thus, the following equation can be used to describe a process that depends on k predictor variables denoted by x_{t1}, x_{t2},…, x_{tk}, and contains a noise term in the form of an ARMA(p,q) model:

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \mu_t \tag{1}$$

where μ_t is an ARMA process with zero mean value. This has the following basic form [36]:

$$a(L)\mu_t = b(L)\varepsilon_t \tag{2}$$

where $t = 1, \ldots, K$$y_t$ is the response series, $\beta_1, \ldots, \beta_k$ are the regression coefficients for k predictor variables, ε_t is the innovation series, $L^j y_t = y_{t-j}$ is the differencing operator, $a(L) = \left(1 - a_1 L - \cdots - a_p L^p\right)$ is the p-degree nonseasonal autoregressive polynomial, and $b(L) = \left(1 + b_1 L + \cdots + b_q L^q\right)$ is the q-degree nonseasonal moving average polynomial. The ARMA process described in Eq. 2 can also include seasonal and nonseasonal integration terms to handle the non-stationarity of the modeled time series, resulting in the SARIMA model for μ_t [36]:

$$a(L)A(L)(1-L)^D(1-L^s)\mu_t = b(L)\varepsilon_t \tag{3}$$

where $A(L) = \left(1 - A_1 L - \cdots - A_{p_s} L^{p_s}\right)$ is the p_s-degree seasonal autoregressive polynomial, $(1-L)^D$ is the D-degree nonseasonal integration term, and $(1-L^s)$ is the s-degree seasonal integration term.

As mentioned previously, if the residuals obtained after the regression in Eq. 1 is applied are expected to exhibit correlation, it is reasonable to perform a least-squares fit to obtain $\hat{\beta}$ and then analyze and model the residuals [35]:

$$\hat{\mu}_t = y_t - \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \cdots + \hat{\beta}_k x_{tk} \tag{4}$$

In this equation, the estimates of the parameter vector $\boldsymbol{\beta}$ for the autocorrelated residuals are

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\boldsymbol{\mu} \tag{5}$$

where $\widehat{\mathbf{V}}$ is the covariance matrix estimate, and $\widehat{\mathbf{V}}$ is now obtained iteratively from the maximum likelihood estimates of the parameters in $\hat{\mu}_t$ by minimizing the log-likelihood objective function $\mathcal{L}$ given by [37]

$$\log\mathcal{L} = -\frac{K}{2}\log(2\pi) - \frac{K}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{K}\varepsilon_t^2 \tag{6}$$

where *K* is the number of samples in the dataset, and σ is the variance of innovations. A SARIMA model is typically defined using a single integer seasonal cycle (defined by the degree p_s in Eq. 3), possibly complemented by seasonal integration (given by the degree *s* in the same equation).

If the Box–Jenkins method is used to identify a forecasting model [35], the time series should be stationary. The non-stationarity of a time series can be heuristically inferred by examining the autocorrelation function; however, a rigorous treatment requires formal tests, specifically, a test for the presence of a unit root in the autoregressive part of the model. In the case of the augmented Dickey–Fuller test used here, this is performed for the (autoregressive) model

$$y_t = y_{t-1} + \beta_1\Delta y_{t-1} + \cdots + \beta_r\Delta y_{t-r} + \varepsilon_t \text{ – null against } y_t = \phi y_{t-1} + \beta_1\Delta y_{t-1} + \cdots + \beta_r\Delta y_{t-r} + \varepsilon_t, \phi < 1 \tag{7}$$

where *r* is the user-defined number of lags. The following commonly used criterion for determining $r_{\max}$ is applied [37]:

$$r_{max} = \left[12.\sqrt[4]{\left(\frac{N}{100}\right)}\right] \tag{8}$$

where [.] is the integer part of the number, and *N* is the number of samples.

3.2 Wavelet transform

The WT is one of the most powerful methods in signal processing, and is a highly effective technique for the analysis of multiscale systems (MS) [38]. It is based on the use of wavelets, which are specialized functions that are localized in both the time and the frequency domain, and allow the separation of slow and fast dynamics of MS [39].

The continuous WT (CWT) is defined as follows [38], [40]:

$$Wx(\tau,s) = \frac{\langle x,\psi_{\tau,s}\rangle}{\|\psi_{\tau,s}\|_2^2} = \int_{-\infty}^{+\infty} x(t)\psi_{\tau,s}^*(t)dt \text{ with } \psi_{\tau,s} = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right),\ \tau,s\in\mathbb{R}\ s\neq 0 \tag{9}$$

where $\psi(t)$ is the "mother wavelet," which satisfies the zero-average and unity-norm conditions

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 = \hat{\psi}(\omega)_{\omega=0} \qquad \|\psi(t)\|_2 = 1 \tag{10}$$

and $\psi_{\tau,s}$ is a wavelet generated by scaling and translating the mother wavelet. The first condition in Eq. 10 implies that wavelets can be regarded as band-pass filters, where scaling moves the center frequency and the bandwidth of the wavelet filter. To address its considerable redundancy, the CWT should be evaluated only at certain scales, and with translations determined by the length of a wavelet at a given scale. Assuming $s = 2^j$ and $\tau = m2^j$ with $j, m \in \mathbb{Z}$, the discrete WT (DWT) corresponds to the evaluation of the CWT at scales s and translations τ [35]:

$$Wf(m,j) = \int_{-\infty}^{+\infty} f(t)\psi^*_{m2^j,2^j}(t)dt \quad \text{and} \quad \psi_{m2^j,2^j}(t) = \frac{1}{2^{j/2}}\psi\left(\frac{t-m2^j}{2^j}\right) \tag{11}$$

Then, by using the definition of the scaling function for the CWT [38], [40]

$$\left|\hat{\phi}(\omega)\right|^2 = \int_1^{\infty} \left|\hat{\psi}(s\omega)^2\right| \frac{ds}{s} = \int_{\omega}^{\infty} \frac{|\hat{\psi}(\xi)^2|}{(\xi)} d\xi \tag{12}$$

it is possible to decompose a signal into an approximation component at given scale j_0 and detail components (up to the finest level of resolution):

$$x(t) = \sum_m a_{m,j_0}\phi_{m,j_0}(t) + \sum_{j\geq j_0}\sum_m d_{m,j_0}\psi_{m,j_0}(t) = A_{J_0} + \sum_{j=1}^{J_0} D_j \tag{13}$$

where $\{a_{m,j_0}\}$ are the approximation coefficients at a given scale, and $\{d_{m,j_0}\}, j = 1, \ldots, j_0$, are the detail coefficients at this level and all finer scales. Thus, the signal can be reconstructed by summing the approximation component at a given scale and all the detail components up to this scale.

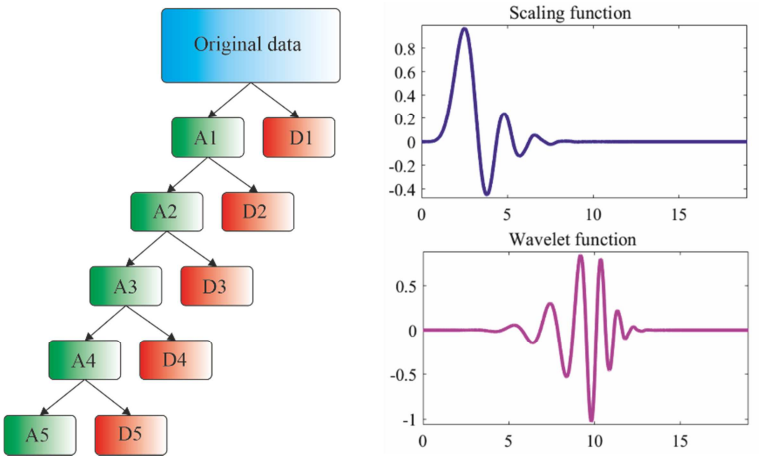


Fig. 5. Wavelet decomposition of original time series to level 5 (left), and scaling and wavelet function for db10 wavelets (right).

In this study, we used Daubechies wavelets of db10 type, with scaling and wavelet functions shown in the right part of Fig. 5. The maximum level of decomposition is given by the number of samples in the given data, which in our case (929 samples for heating seasons) corresponded to a value of 10. Determining the optimal level of decomposition remains an open question in time-series forecasting. We experimentally demonstrated that a level of 5 achieved a reasonable balance between decomposition complexity and improvement over the reg(S)ARMA approach. As shown in the left part of Fig. 5, the original signal can be reconstructed by summing the highest-level approximation component (A5) and all details up to this level (D1 + D2 + D3 + D4 + D5). Each detail signal at the *j*-th level allows the observation of cyclic components with periods between 2^{j-1} and 2^{j} [41]; thus, details D5 and d4 contain components with periods between 16 and 32 days, and 8 and 16 days, respectively, which can be recognized visually. Likewise, the remaining details contain cyclic components with periods between 4 and 8 days, 2 and 4 days, and 1 and 2 days (Fig. 6).

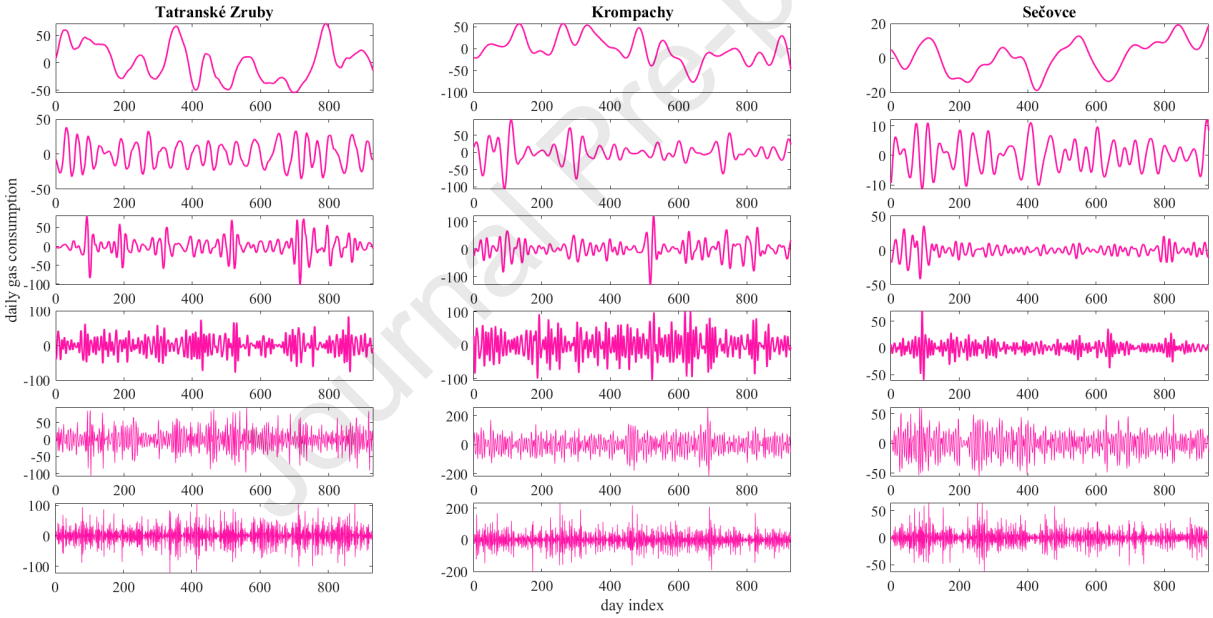


Fig. 6. Approximation and detail components of entire time series characterizing daily gas consumption (929 days).

3.3 Forecasting using neural networks with genetic-algorithm-optimized regressors

Generally, two basic strategies are used for multistep-ahead forecasting: direct and recursive [42]. The former is computationally more demanding, and in general results in quite different models for each forecasting horizon [43]. In contrast, the latter uses one model for different forecasting horizons and is based on minimizing the error for one-step-ahead forecasts. Owing to its compactness and lower computational complexity, the latter appears to be more widely used in typical forecasting frameworks.

If, for simplicity, we assume that there are *P* samples of the univariate process $\{y_1, \dots, y_P\}$ from, possibly, a NAR process, then this can be expressed as [43]

$$y_t = \varphi(\mathbf{x}_{t-1}) + \varepsilon_t \quad \text{and} \quad \mathbf{x}_t = [y_t, \dots, y_{t-m+1}]' \tag{14}$$

where ε_t is an iid process with zero mean and σ^2 variance. In fact, for one-step-ahead forecasting, we estimate the model

$$y_t = \psi(\mathbf{x}_{t-1}; \boldsymbol{\theta}) + e_t \quad \text{and} \quad \mathbf{x}_t = [y_t, \dots, y_{t-n+1}]' \tag{15}$$

where ψ is a function for the estimated model, $\boldsymbol{\theta}$ is the model parameter vector, and *n* is the embedding dimension. Usually, the mean-squared-error form of the objective function is used in the optimization, that is, $E\left[\left(y_{t+1} - \psi(\mathbf{x}_t; \boldsymbol{\theta})\right)^2 | \mathbf{x}_t\right]$, with parameters estimated as follows [43]:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_t \left(y_t - \psi(\mathbf{x}_{t-1}; \boldsymbol{\theta})\right)^2 \tag{16}$$

The recursive strategy for *m*-step-ahead forecasting can be computed as

$$\hat{\psi}^{(m)}(\mathbf{x}_t) = \hat{\psi}\left(\left[\hat{\psi}^{(m-1)}(\mathbf{x}_t), \dots, \hat{\psi}^{(m-p)}(\mathbf{x}_t)\right]'\right) \tag{17}$$

where $m > 0$, $\hat{\psi}$ is the function ψ but with the estimated parameter vector $\widehat{\boldsymbol{\theta}}$ in place of the parameter vector, and *p* is the estimated embedding dimension.

The form of Eq. 17 is general, and $\hat{\psi}$ can be approximated using a linear or a nonlinear model. Neural networks represent highly flexible models capable of approximating complicated nonlinear functions, provided that their structure contains a sufficient number of neurons in the hidden layer [44]. When a sigmoid activation function is used, the output of an NN consisting of *n* neurons in the hidden layer can be expressed as [45]

$$v(u) = \sum_{k=1}^{n} \frac{a_k}{e^{\left(b_k(u - c_k)\right)} + 1} \tag{18}$$

Substituting this into Eq. 17 yields the following formula for computing *h*-step-ahead forecasts using a sigmoid NN:

$$\hat{\psi}^{(m)}(\mathbf{x}_t) = \sum_{k=1}^{n} \frac{a_k}{e^{\left(b_k(\mathbf{x}_t - c_k)\right)} + 1} \left(\left[\left(\sum_{k=1}^{n} \frac{a_k}{e^{\left(b_k(\mathbf{x}_t - c_k)\right)} + 1}\right)^{(m-1)}, \dots, \left(\sum_{k=1}^{n} \frac{a_k}{e^{\left(b_k(\mathbf{x}_t - c_k)\right)} + 1}\right)^{(m-p)}\right]'\right) \tag{19}$$

It should be noted that the proposed model consisted of several NNs, with their number depending on the level of the WT decomposition. Accordingly, the correspondence between the estimated embedding dimension *p* of

the process model and embedding dimension of the WT-component models was evidently lost. To address this, the regressor vector of each model was obtained using a search algorithm. Assuming that the presence or absence of a given regressor can be expressed using binary logic, the use of a binary GA appears to be a suitable solution for this problem [46].

3.3.1 Binary genetic algorithm

A binary GA can be considered a special case of an integer (or mixed-integer) GA, where the possible values for an individual are restricted to 0 and 1. This implementation uses Laplace crossover and power mutation together with a tournament function for the selection of individuals [47]. Initially, a population of individuals is randomly generated; this population forms an initial pool of possible solutions (Fig. 7). If none of the defined stopping criteria is satisfied, a mating pool is generated using the tournament selection function. The genetic information contained in the individuals in the mating pool is then recombined as follows: Let $a^1 = (a_1^1, a_2^1, \dots, a_n^1)$ and $a^1 = (a_1^2, a_2^2, \dots, a_n^2)$ denote the parents, and let $b^1 = (b_1^1, b_2^1, \dots, b_n^1)$ and $b^1 = (b_1^2, b_2^2, \dots, b_n^2)$ be their children generated using the random numbers $\varphi_i, \psi_i \in [0,1]$ and the Laplace distribution (LD) [47]:

$$\gamma_i = \begin{cases} \alpha - \beta \log(\varphi_i), \psi_i \leq 0.5; \\ \alpha + \beta \log(\varphi_i), \psi_i > 0.5, \end{cases} \tag{20}$$

where γ_i is a random number following the LD, α is the location parameter, and $\beta > 0$ is the scaling parameter. The scaling parameter should be an integer in the case of a binary GA implementation. The children are then generated according to

$$b_i^1 = a_i^1 + \gamma_i |a_i^1 - a_i^2| \qquad b_i^2 = a_i^2 + \gamma_i |a_i^1 - a_i^2| \tag{21}$$

To improve the exploration capability of the algorithm, new genetic information should be introduced using a mutation operator, which is based on the power distribution [47]. Let δ denote a random number following the power distribution, δ_1 a uniform random number between 0 and 1, and m the mutation index (integer in this case). Then,

$$\delta = (\delta_1)^m, \quad a = \begin{cases} \bar{a} - \delta(\bar{a} - a^l), \ \omega < \psi \\ \bar{a} + \delta(a^u - \bar{a}), \ \omega \geq \psi, \end{cases} \qquad \omega = \frac{a - a^l}{a^u - a} \tag{22}$$

where a is the mutated solution, $\bar{a}$ is the parent solution, ψ is a uniform random number between 0 and 1, and a^l and a^u are the lower and the upper bound, respectively, of the decision variables.

After the crossover and mutation operations are performed, it is important to ensure that the solutions are integer numbers; this is easily carried out as follows. If the solution is not already an integer, it is set equal to the closest (higher or lower) integer with a probability of 0.5. Constraint handling is performed by using the penalty function defined in [47], where infeasible solutions depend on the amount of constraint violation but also on individuals in the population. The algorithm is summarized in the flowchart shown in Fig. 7.

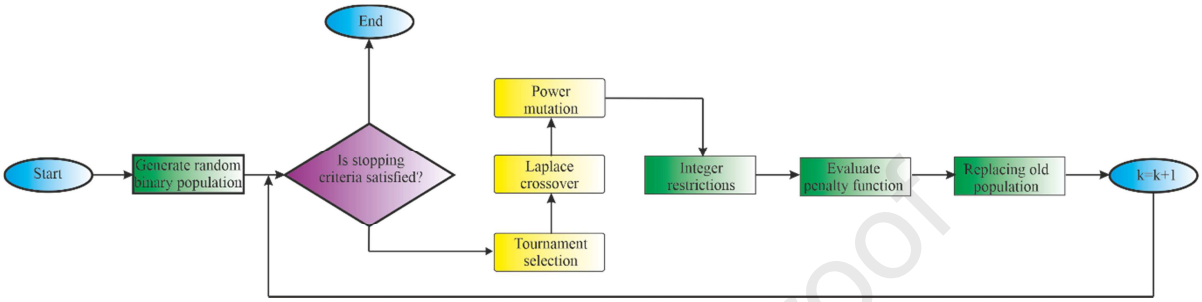


Fig. 7. Flowchart of binary GA with Laplace crossover and power mutation.

3.3.2 Methodology for identification and use of the proposed model

As using the GA-optimized regWANN model requires several methods, a step-by-step methodology for its use is now introduced. It comprises model derivation (identification) and application to week-ahead gas consumption forecasting.

Identification

It is assumed that in this part, sufficient data for the development of the forecasting model are available. Even though it is difficult to set a definite number for the necessary number of samples, we used more than 800 samples. Each sample corresponds to the total daily gas consumption in a building. Only heating seasons are considered, and both the start and the end of the seasons can be determined based on the specific conditions for any building. In addition to the daily gas consumption data, the recorded average outdoor temperature for the given days should be available so that the temperature regression model can be identified. The performance of the regWANN model after identification should be evaluated using an independent test dataset. The step-by-step procedure for the identification of the GA-optimized regWANN model is as follows (Fig. 8 top):

1. Prepare the training dataset containing the daily gas consumption in a given building for the period of interest. It is assumed that only heating seasons are considered.

2. Determine the best linear fit of $Q_{td}^k = f(v^k)$, where $k = 1, \dots, N$ is the number of days in the entire dataset, Q_{td}^k is the total gas consumption on the k-th day, and v^k is the average outdoor temperature on the k-th day (Fig. 11).
3. Calculate the gas consumption forecasts using the linear regression model $(\hat{Q}_{tdr}^k)$ from the previous step and obtain the residual time series (R_{td}^k) by subtracting $\hat{Q}_{tdr}^k$ from the recorded gas consumption Q_{td}^k.
4. Perform wavelet decomposition of R_{td}^k to the desired level of resolution.
5. Select the structure, training algorithm, and hyperparameters of each NN model intended for residual wavelet component forecasting.
6. The NNs from the previous step are used as NAR models, where the lags in the regressor vector should be determined. Select the parameters for the binary GA, where the position in each chromosome vector corresponds to the presence or absence of a given lag in the regressor vector (Fig. 10).
7. The fitness function is defined so that a given error metric (e.g., NRMSE) is minimized. To reduce model complexity, include a term for the size of a regressor vector (Eq. 24). The error metric should be evaluated on a test dataset.

Forecasting

In the forecasting part, the historical data of the daily gas consumption are required for each NAR model. The size of the required past data depends on the lags present in the optimized regressor vectors (Table. 5). As the models obtained from the identification part use the wavelet decomposition of the temperature regression residuals, the historical data necessary for forecasting are the past time series of the residuals R_{td}^{k-p}, where $p = 1, \dots, P$, and P is the maximum lag in any optimized regressor vector. The step-by-step forecasting procedure using the GA-optimized regWANN model is as follows (Fig. 8 bottom):

1. Calculate the week-ahead gas consumption point forecast (R_{td}^7) using the linear regression model from step 2 of the identification procedure and the week-ahead temperature forecast for a given location.
2. Calculate past gas consumption forecasts using the linear regression model $(\hat{Q}_{tdr}^{k-p})$ and either the recorded (v^{k-p}) or forecasted $(\hat{v}^{k-p})$ daily average temperatures, and obtain the residual time series (R_{td}^{k-p}) by subtracting $\hat{Q}_{tdr}^{k-p}$ from the recorded gas consumption Q_{td}^{k-p}.
3. Perform the wavelet decomposition of R_{td}^{k-p} to the resolution level selected in the identification part.

4. Use the set of GA-optimized NAR models from the identification part to obtain the week-ahead approximation and detail forecasts (a_L^7 and d_l^7, where l = 1,…, L, and L is the wavelet decomposition level) based on the R_{td}^{k-p} as input to the NAR models.
5. Correct R_{td}^7 using the value of $a_L^7 + \sum_{l=1}^{L} d_l^7$.

We present the implementation of this methodology for the buildings in question in the following subsection.

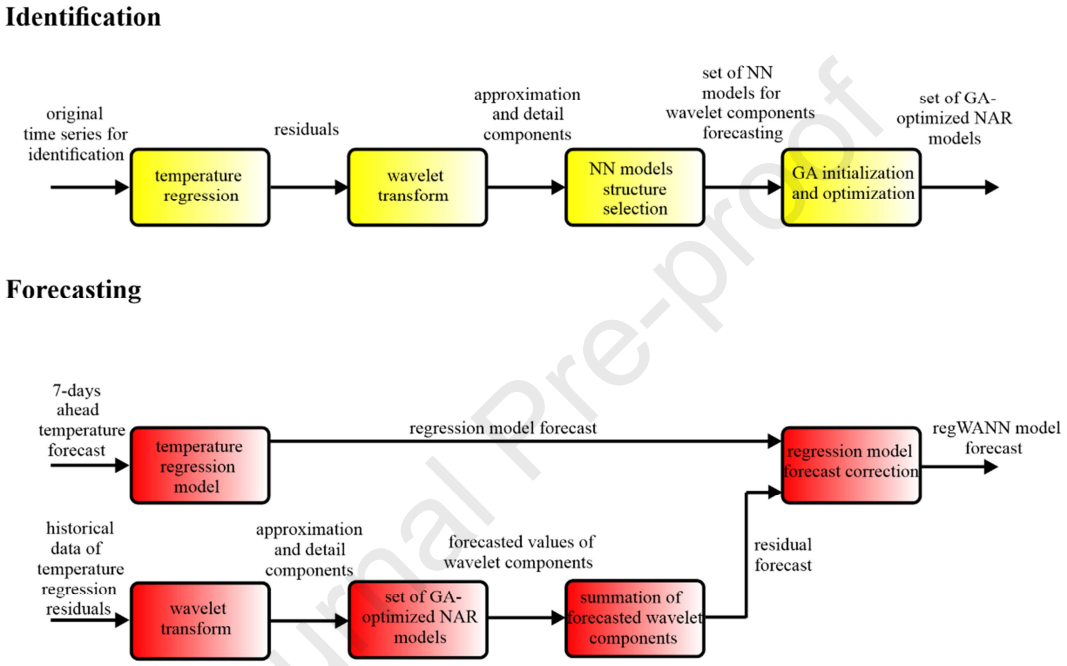


Fig. 8. Scheme of improved GA-optimized regWANN model.

3.3.2.1 Implementation of the proposed methodology for week-ahead forecasts

By combining the methods described above, the GA-optimized regWANN model was used for week-ahead gas consumption forecasting. The structure of this model is shown in Fig. 9, specifically, the combination of the model preparation steps as well as the six NAR models for forecasting the wavelet components. The temperature regression step was performed using the models in Table 2, and only the residuals were decomposed into one approximation component (A5) and five detail components (D1−D5) using the DWT. The inputs to the NAR models were regressor vectors containing a specific number of lags used to forecast a given wavelet component.

The number of lags in a given regressor vector was determined using the binary GA, as shown schematically in Fig. 10.

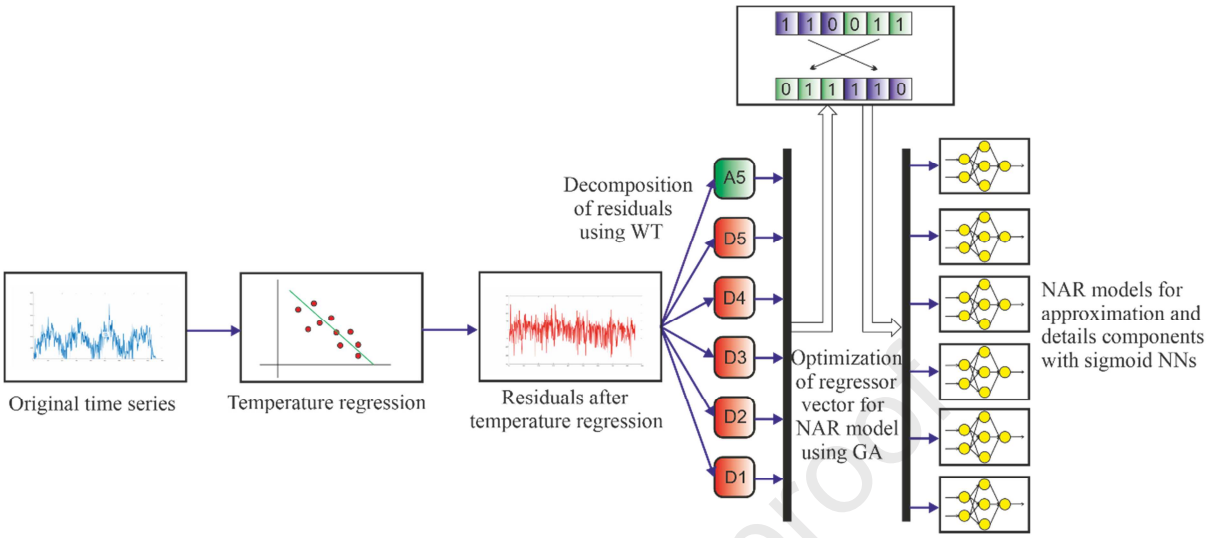


Fig. 9. Scheme of improved GA-optimized regWANN model.

Week-ahead forecasting using the GA-optimized regWANN model is shown separately in Fig. 10. This structure corresponds to each of the NAR models shown in Fig. 9, where the number of inputs is equal to the number of nonlinear regressors (represented by 1s in the solution vector) determined using the binary GA. The model uses a recursive strategy (Eq. 17) to forecast h steps ahead (h = 7 for week-ahead forecasts), regarded as feedback of the forecasted output $\hat{y}(k+h)$. To train the sigmoid NNs in the NAR models, a combination of four iterative search methods was used: subspace Gauss–Newton least squared search, the adaptive version of this method, the Levenberg–Marquardt algorithm, and steepest descent least squares search. The selection of a particular algorithm was based on the direction that reduced the estimation cost first [45]. The number of neurons in each NAR model was set to 10.

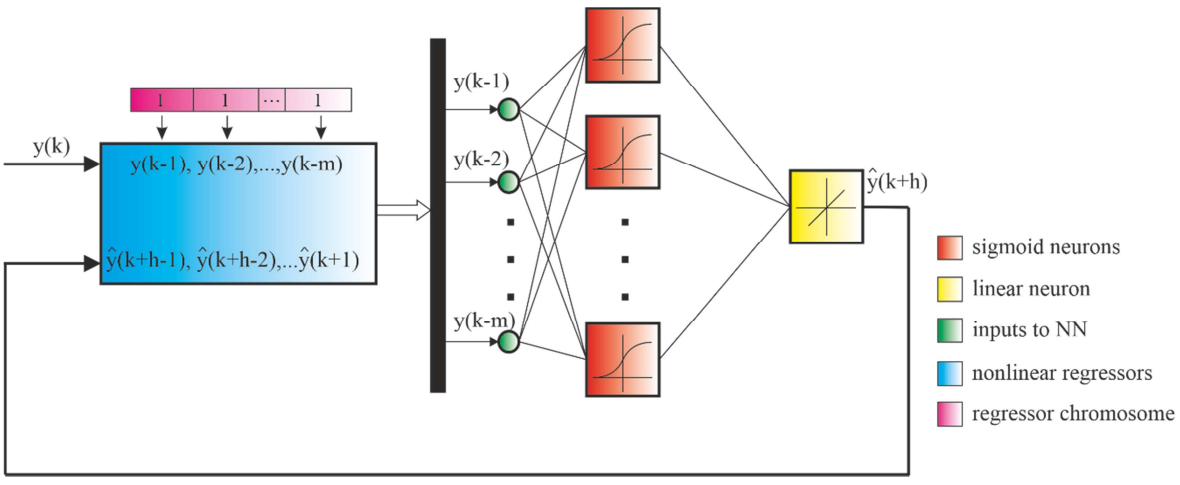


Fig. 10. Seven-day-ahead forecasting using the sigmoid-NN-based NAR model.

To compare the forecasting performance of the given models, two error criteria were selected: the mean absolute error (MAE) and the NRMSE expressed in percent of fit. The MAE criterion was deemed suitable because it provides a rapid assessment of the mean forecasting error expressed in the units of the variable of interest (in our case, cubic meters), whereas NRMSE is an error measure typically used in the identification process to express the goodness of fit (100% implies that the model perfectly fits the data). The formulas for these indices are as follows:

$$MAE = \frac{1}{N}\sum_{k=1}^{N}|y(k) - \hat{y}(k)| \qquad NRMSE = \left(1 - \frac{\sqrt{\sum_{k=1}^{N}(y(k)-\hat{y}(k))^2}}{\sqrt{\sum_{k=1}^{N}\left(y(k)-\frac{1}{N}\sum_{k=1}^{N}y(k)\right)^2}}\right)\times 100\% \tag{23}$$

where $y(k)$ is the actual value at the *k*-th time instant, $\hat{y}(k)$ is the forecasted value at the *k*-th time instant, and *N* is the number of samples.

To apply the GA algorithm successfully, it was necessary to define an appropriate fitness function so that individual solutions could be compared. In the definition of the fitness function, two factors were considered: forecasting error and model complexity. As model performance on unseen data is particularly important, the NRMSE (Eq. 24) on the test dataset was included in the fitness function. In addition, to favor less complex models among models with similar forecasting performance, the regressor vector size was also considered in the fitness function, which was defined as

$$F = -\left(K_e \times NRMSE + K_r \times \frac{K}{|\{x \in \boldsymbol{\mathcal{R}} | x = 1\}|}\right) \tag{24}$$

where K_e is the error weight constant (set to 0.5), K_r is the regressor size (RS) weight constant (set to 0.5), and *K* is the normalizing constant (set to 300). The second term in Eq. 24 corresponds to the number of 1s in the regressor vector $\boldsymbol{\mathcal{R}}$.

The parameter values used in the GA for regressor-vector optimization are summarized in Table 1.

Number of neurons in NAR models	Number of generations	Number of individuals in population	Number of tournament individuals	Number of elite individuals	Number of GA runs for one case
10	30	20	4	2	10

Table 1. Settings used for GA optimization of regressor vector in regWANN models.

4 Results and discussion

4.1 Temperature regression model

Initially, temperature regression was performed for the gas consumption time-series to obtain residuals, which were themselves modeled using different approaches. The regression part of the reg(S)ARMA model as well as of the GA-optimized regWANN model was based on the relationship between the daily average outdoor temperature and the daily gas consumption in the given buildings. This relationship was derived only for the extracted data corresponding to the four heating seasons shown in Fig. 11. Visual inspection of these graphs indicates that there is a strongest correlation between temperature and daily gas consumption in the hotel in Tatranské Zruby, whereas the weakest correlation is observed in the health care building in Sečovce. Using the results in Table 2, this correlation was confirmed by the value of R^2, which was almost 15% higher for Tatranské Zruby (0.7382) than for Sečovce (0.6319) and Krompachy (0.6646) (the latter two exhibited roughly similar correlation). The magenta lines in Fig. 11 show the best fits using a linear model for the data from all heating seasons. The uncertainty associated with these models can be evaluated using the calculated 95% parameter bounds, which have the lowest relative width for the Tatranské Zruby model. The gas consumption profile in Tatranské Zruby differs from those in both Krompachy and Sečovce owing to the heating system, which was never fully switched off. This is in contrast with Krompachy and Sečovce, where for a relatively large number of days, the total daily gas consumption was zero (represented by the data points lying on the x-axis). The difference can be attributed to both the purpose of the building in Tatranské Zruby (a hotel is expected to have a certain number of guests throughout the entire year) and to its northernmost geographical location.

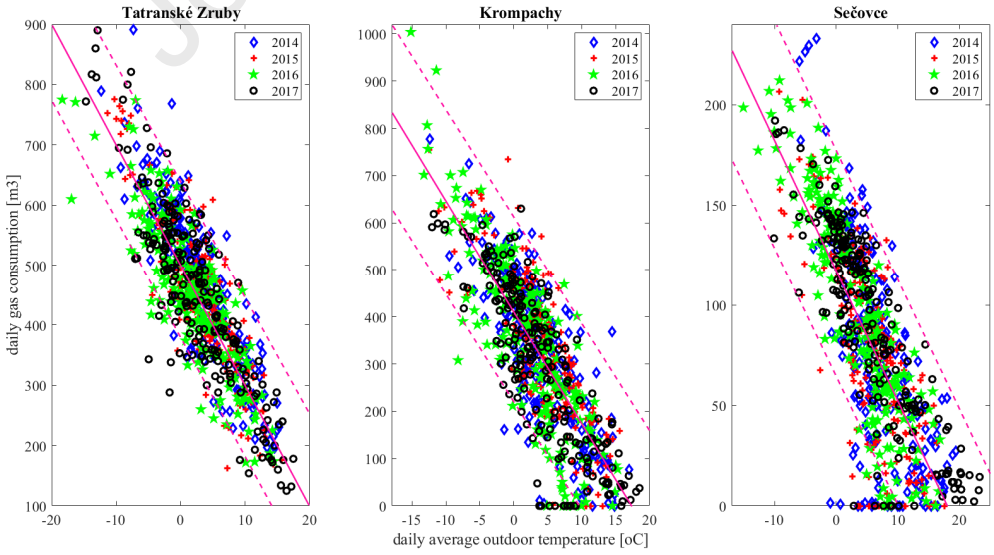


Fig. 11. Best linear fits and 95% prediction bounds of gas consumption-outdoor temperature relationship used as the temperature regression model

4.2 Performance of ARMA/SARMA models

As the residuals after temperature regression were serially correlated, it was useful to analyze their ACF and PACF to infer suitable forecasting models. The ACF and PACF plots are shown in Fig. 12 for the default number of lags (20), which is considered sufficient for several types of time series [35]. The plots of both types indicate similar correlations for the Krompachy and Sečovce buildings, with strong weekly seasonality indicated by high values at lags 7 and 14. This implies the possible suitability of SAR(I)MA models with a seasonal period of 7 for modeling the residuals in both cases. In contrast, the ACF and PACF for Tatranské Zruby do not clearly indicate a strongly seasonal character, and the shape of the plots points to the AR(I)MA model (PACF drops rapidly after the first lag, but the correlations at several subsequent lags appear to be sufficiently significant to consider only the AR model). The stationarity of the residual time series was tested using the augmented Dickey–Fuller test for the presence of a unit root, with the number of lags determined using the procedure described in [37], and r_{max} calculated using Eq. 8 for the merged datasets of all four years. As the number of samples in the datasets was the same (853), r_{max} calculated using Eq. 8 was 20 for each building. The absolute values of the test statistic were larger than 1.6 in all three cases (Table 2); therefore, this value of r_{max} was used to obtain the results of the ADF test. As shown in the table, the null hypothesis was rejected at the 0.05 significance level, indicating sufficient evidence for favoring the alternative hypothesis (i.e., that of the absence of a unit root). Consequently, ARMA was selected for modeling the residuals of the Tatranské Zruby time series, and SARMA for both Krompachy and Sečovce.

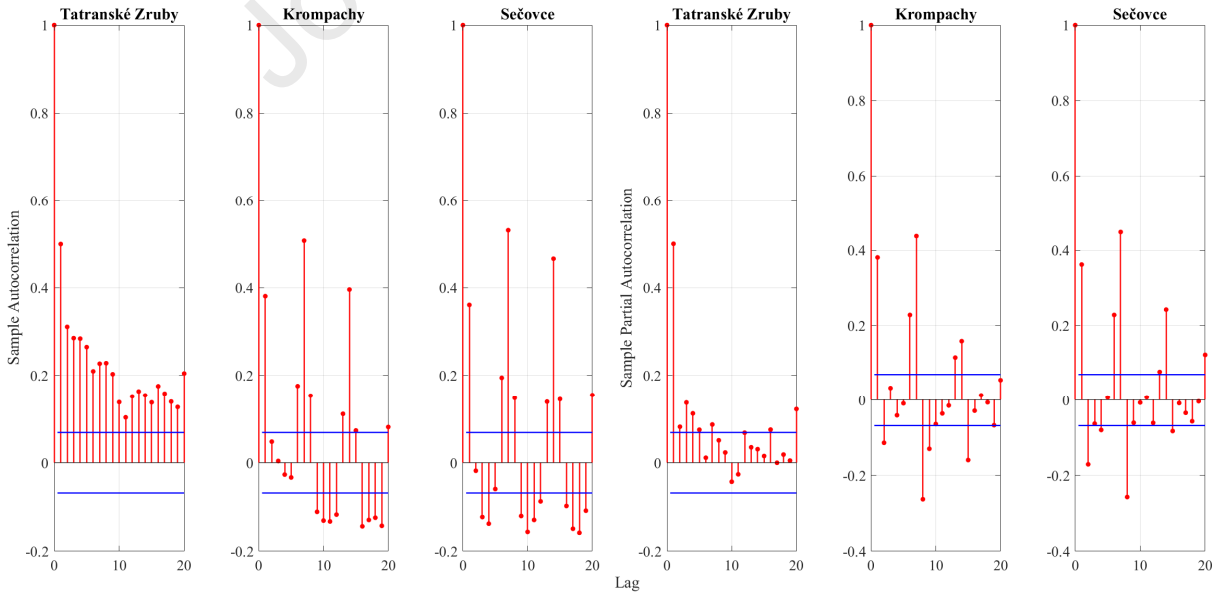


Fig. 12. Autocorrelation (left three) and partial autocorrelation (right three) functions of temperature regression model residuals

Fit	T. Zruby	Krompachy	Sečovce	**ADF test**	TZ	K	S
Parameters and 95% parameter bounds	$p_1 = -19.38$ $(-20.16, -18.61)$ $p_2 = 507.6$ $(502.6, 512.5)$	$p_1 = -23.96$ $(-25.11, -22.82)$ $p_2 = 412$ $(404.2, 419.8)$	$p_1 = -6.589$ $(-6.927, -6.251)$ $p_2 = 120.9$ $(118.2, 123.6)$	*p*-value	0.001	0.001	0.001
				test statistic	−3.5077	−5.0739	−3.6561
SSE	3.77×10^6	8.257×10^6	8.007×10^5	critical value	−1.9414	−1.9414	−1.9414
R^2	0.7382	0.6646	0.6319	significance	0.05	0.05	0.05
$Adj.R^2$	0.7379	0.6642	0.6314	lags	20	20	20
RMSE	66.55	98.5	30.67	H0 reject	true	true	True

Table 2. Fit results for temperature regression models and unit root test results using the augmented Dickey–Fuller method.

The models considered for forecasting were evaluated based on their parsimony expressed through the AIC and BIC criteria as well as their ability to pass the whiteness test for their residuals. The models shown in Table 3 were selected from a group of models with various orders of autoregressive and moving-average terms. For Krompachy and Sečovce, the models with structure SARMA$(1,0,0)\times(1,0,1)_7$ had the lowest AIC and BIC values, and passed the whiteness test for the residuals. A period of seven days was selected based on the analysis of ACF and PACF shown in Fig. 12. Considerably lower variance and values of the information criteria for the Sečovce model can be associated with relatively lower average values of gas consumption. Lower-order models for the Tatranské Zruby dataset were unable to pass the whiteness test, and its final order (ARMA(4,0,3)) was obtained by successively increasing the orders and evaluating the AIC/BIC values. Further increasing the orders did not result in better model parsimony, and thus the first model to pass the test was selected. Figure 13 shows the graphic residual analysis of all three models using histograms, as well as the ACF and QQ (quantile–quantile) plots of the residuals. The ACF plots of the residuals for all models indicate that they are virtually uncorrelated at the 0.05 significance level for the first 20 lags. The QQ plots indicate larger deviations of the residuals from the normal distribution at the ends, suggesting the suitability of the t-distribution, which can handle heavy tails. As observed in the histograms, these deviations are due to the presence of outliers, which are clearly more probable to occur than in the case of the normal distribution. The distribution of the residuals is closer to normal in the Tatranské Zruby model, which has the highest number of degrees of freedom (approximately 10), compared to 3.4 and almost 5 in the Krompachy and Sečovce models, respectively. The residual outliers are related to extreme values in the temperature regression model errors, the number of which was larger in the Sečovce and Krompachy models. Moreover, the distribution of residuals in the Krompachy model is slightly skewed to the right, resulting in the smallest number of degrees of freedom for the t-distribution.

Place	Model	Equation	AIC	BIC	Var	DoF
TZ	ARMA(4,0,3)	$(1-1.131L+1.125L^2-1.000L^3+0.145L^4)y_t = \\ =(1-0.737L+0.883L^2-0.645L^3)\varepsilon_t$	9299.0	9336.9	3117.6	9.986
K	SARMA(1,0,0) ×(1,0,1)$_7$	$(1-0.453L)(1-0.927L^7)y_t = \\ =(1-0.641L^7)\varepsilon_t$	9729.4	9748.3	5212.4	3.366
S	SARMA(1,0,0) ×(1,0,1)$_7$	$(1-0.390L)(1-0.925L^7)y_t = \\ =(1-0.615L^7)\varepsilon_t$	7722.3	7741.2	495.6	4.916

Table 3. Results of identification of forecast models for gas consumption in the reg(S)ARMA approach

The models derived above were used for one-week-ahead gas consumption forecasts for each building, and the results are shown in Fig. 14. These are *ex post* seven-day-ahead forecasts during the 50-day period of the test data (Fig. 2) with recorded (i.e., not forecasted) temperatures. In Fig. 14, a comparison is also made between the forecasts by pure linear regression models (brown/red line) and reg(S)ARMA error models (brown/yellow line).

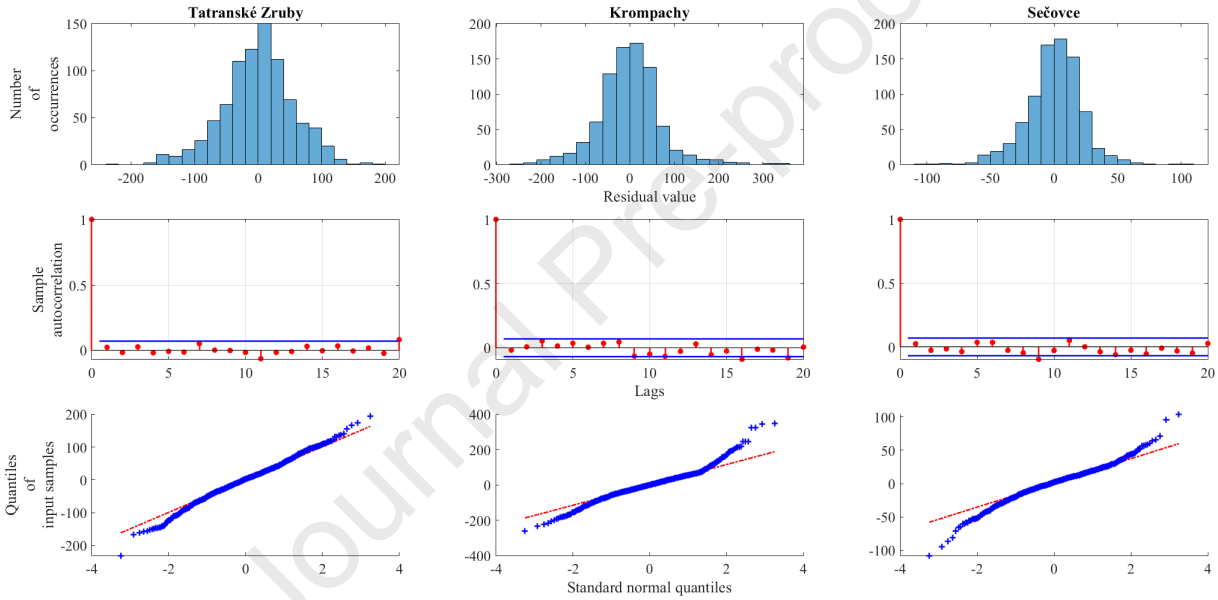


Fig. 13. Histograms, autocorrelation, and QQ plots of residuals for (S)ARMA models.

The errors shown in the bottom part of the figure correspond to the differences between the recorded and forecasted consumption values using reg(S)ARMA models. These models allowed reducing the error (MAE) in one-week-ahead forecasting from 50.73 m^3 to 37.65 m^3 for Krompachy, and from 20.74 m^3 to 16.97 m^3 for Sečovce. Interestingly, in the Tatranské Zruby dataset, compared with pure temperature regression, the regARMA model failed to increase forecasting accuracy, as the MAE increased from 45.55 m^3 to 48.38 m^3. In this case, the linear correlation between gas consumption and daily average outdoor temperature was strongest (r^2 = 0.74), and thus the random component of the residuals after regression was more significant than in the other two buildings. This, combined with a lack of obvious weekly seasonality, probably resulted in worse seven-day-ahead forecasting performance for the regARMA model than for a simple temperature regression model.

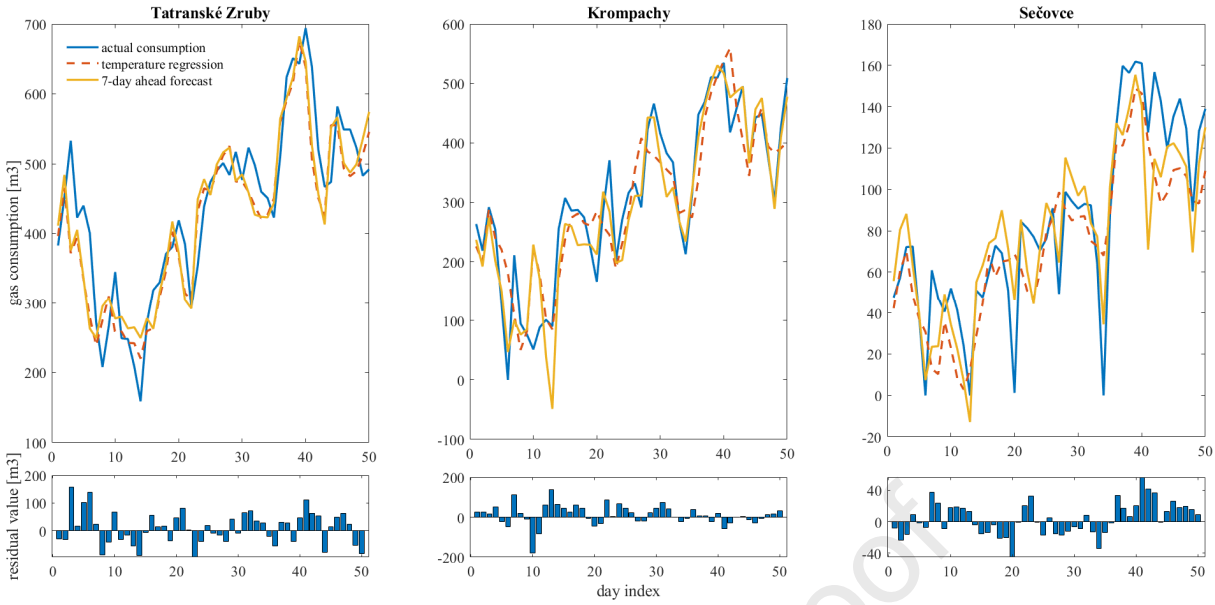


Fig. 14. Seven-day-ahead forecast of daily gas consumption using temperature regression model and ARMA/SARMA error models tested on the 50-day test dataset and recorded temperatures.

In an online scenario, seven-day-ahead forecasted temperatures would be required with reg(S)ARMA or WANN models. As historic data for temperature forecasts were not available for the same periods as for gas consumption forecasts, it was not possible to perform temperature forecasts for the entire test dataset period (50 days). However, seven-day-ahead temperature forecasts were recorded from January 21, 2020, which could be used to test the models by analogy with an online forecasting scenario. In view of the availability of forecasted temperature data, the testing period was set to be from December 1, 2019, to March 10, 2020, containing 101 days in total. Of these, 88 days were used as pre-sampling data for week-ahead forecasts performed for seven days, starting on March 4, 2020, and ending on March 10, 2020. The residuals after the temperature regression that were used as data during the given period are shown in Fig. 15 (left), where three distinct sections are indicated by I, II, and III. The first corresponds to a period for which the residuals were obtained as the difference between the recorded gas consumption and the gas consumption estimated using the linear models from Table 2 based on the recorded temperatures (first 58 days). However, it was demonstrated that forecasting performance was improved when the residuals obtained from the forecasted temperatures were used as pre-sampling data.

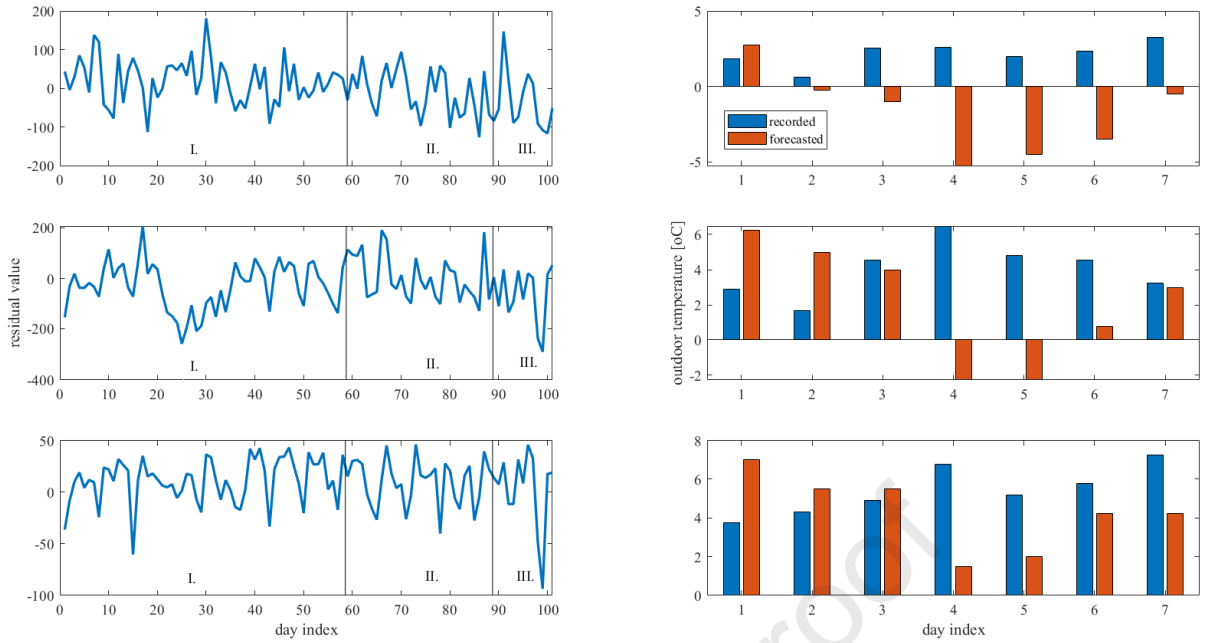


Fig. 15. Recorded and forecasted temperatures for testing period from March 4, 2020, until March 10, 2020, (right column) and presampling residuals used in forecasting models (from December 1, 2019, to March 10, 2020) (left column).

Therefore, starting from January 28, 2020, the residuals obtained as the difference between the recorded gas consumption and gas consumption estimated using the week-ahead forecasted temperatures were used (sections II and III). Section III corresponds to the period for which full forecasts (forecasted temperature + forecasted residuals) were performed. The accuracy of temperature forecasts using data from the Norwegian Meteorological Institute [48] can be evaluated using Fig. 15 (right). The mean absolute error for temperature forecasts in section III was 4.17°C (Tatranské Zruby), 3.86°C (Krompachy), and 2.58°C (Sečovce), whereas for sections II and III, it was 2.92°C, 2.39°C, and 1.93°C, respectively. The reduced accuracy of the temperature forecasts in section III was caused by the forecasted sudden temperature drop in the middle of this period that did not, in fact, occur. The *ex ante* week-ahead forecasting performance of the reg(S)ARMA models can be seen in Fig. 16. As in Fig. 14, a comparison can be made between simple forecasts using the temperature regression models only. Clearly, incorrectly forecasted sudden changes in temperature have profound negative effects on the accuracy of gas consumption forecasts. This is evident in the gas consumption profiles of Krompachy and Sečovce, where the temperature in fact increased in the middle of section III, causing the actual consumption to drop (instead of rising, as forecasted by the models). As a result, the errors in this part of Section III reached approximately 100 m^3 and 200 m^3 for Tatranské Zruby and Krompachy, respectively, and above 50 m^3 for Sečovce. The seven-day MAEs were: 63.21 m^3 (Tatranské Zruby), 94.90 m^3 (Krompachy), and 33.70 m^3 (Sečovce).

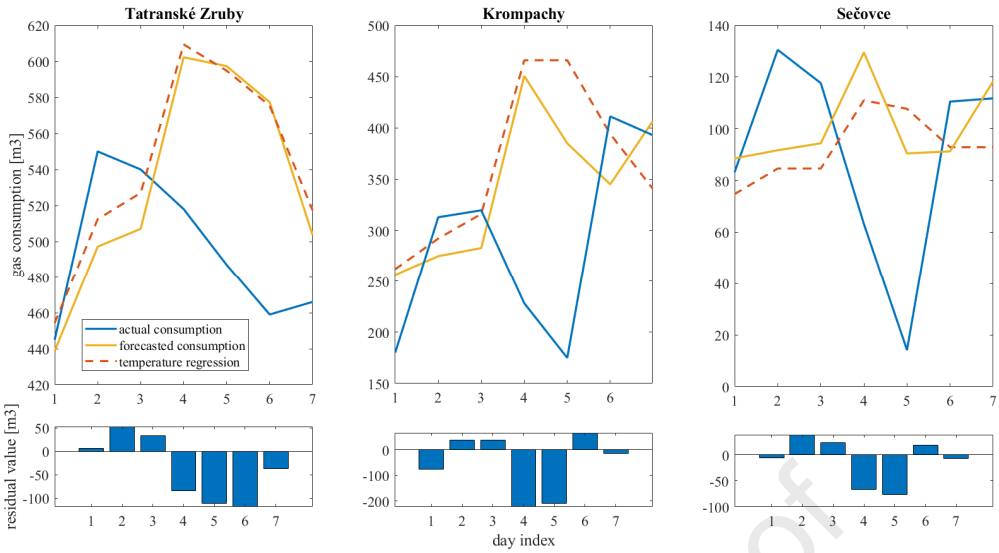


Fig. 16. Seven-day-ahead forecast of daily gas consumption using temperature regression model and ARMA/SARMA error models tested on 7-day test dataset and forecasted temperatures.

4.3 Performance of GA-optimized WANN model

The GA algorithm was run 10 times for regressor optimization of each component so that the average results could be evaluated. Its parameters were determined experimentally using a number of algorithm runs with different parameter settings and resulting performance. As indicated in Table 1, the number of individuals in the population was set to 20, with elitism set to two individuals. Under the given conditions, the effective number of generations for obtaining useful results was relatively low and was set to 30. The results of 10 runs of the GA for each component are shown in Fig. 17, where the points correspond to the mean value of the fitness score, and the width of the error bars indicates the range of one standard deviation around this value. According to Eq. 24, the fitness function consists of an error part (fit value of the model on test data) and an RS part (the number of 1s in the optimized regressor vector); the minus sign is used so that a minimization problem may be formulated. Accordingly, the values on the y-axis in Fig. 17 are fitness scores obtained from each GA run, with each fitness function component weighted equally. The gradual increase in the fitness values with higher-resolution wavelet components observed in the graphs in Fig. 17 can be explained by the lower predictability of these components (and thus lower error part of the fitness value), as well as the higher number of regressors in the regressor vector required for forecasting. In addition, the larger width of standard deviation ranges for the first three components (A5, D5, and D4) can be attributed to the greater sensitivity of the fitness value to the ratio of the NRMSE value and the number of regressors in the regressor vector. The mean and standard deviation values were comparable for each building, except for the A5 and D5 components for Sečovce, where a higher number of small regressor

vectors in the results contributed to the lower fitness values (−387.36 (A5) and −260.76 (D5) compared with −287.00 (A5) and −206.57 (D5) for Krompachy, and −295.55 (A5), −228.55 (D5) for Tatranské Zruby).

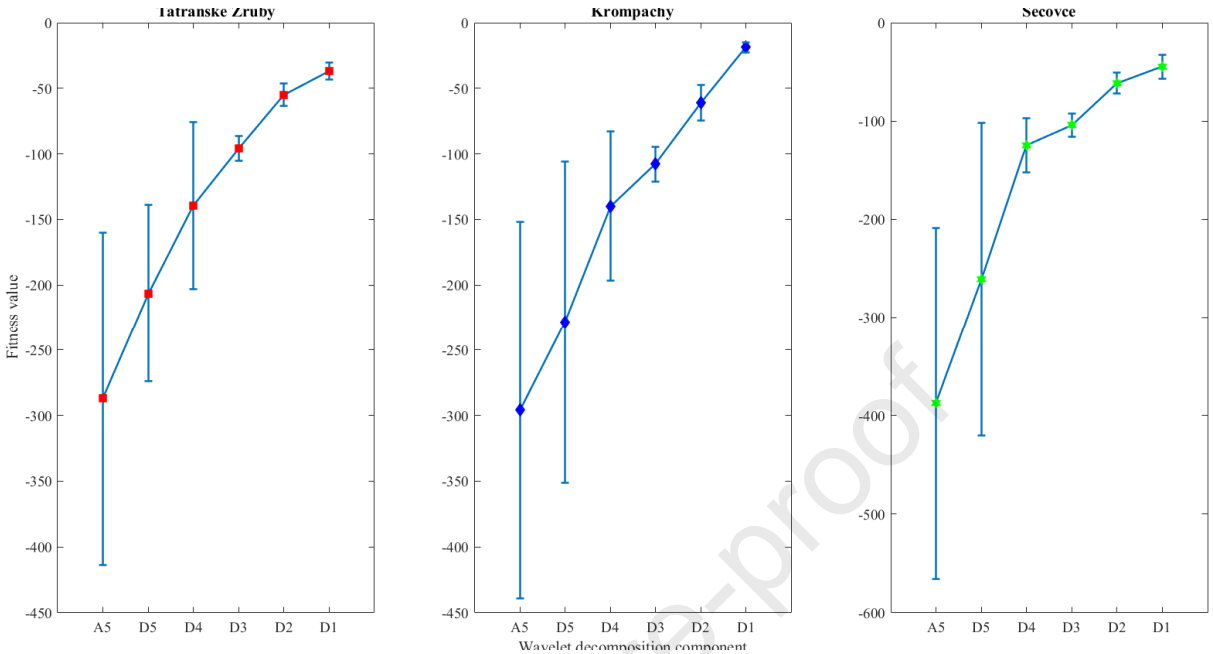


Fig. 17. Error bars of GA results for each of the wavelet decomposition components obtained from 10 runs of the algorithm.

To better evaluate the performance of the GA-optimized models in terms of their forecasting ability, Table 4 summarizes the results using NRMSE, MAE, and the RS indicator. The NRMSE criterion is given in the standard model fit form, expressing the deviation between the recorded data and the model response in percentage error. By contrast, the other criterion directly provides the MAE in units of gas consumption (i.e., cubic meters). The RS indicator is the number of regressors in a regressor vector after GA optimization (integer number). We note the relatively low standard deviation of model fit (NRMSE criterion) when the size of a regressor vector changes, implying that simpler models with a smaller number of regressors could be preferred without significantly worsening forecasting performance. As the standard deviation of the NRMSE criterion increased with the level of wavelet decomposition, it is evident that RS had a dominating effect on the large width of the standard deviation of fitness in the first three components in Fig. 17. On average, the first four components (A5, D5, D4, and D3) could, in all cases, be forecasted seven days ahead with high accuracy: the average value of model fit was under 90% in only one case (D4 for Krompachy). This is severely reduced for the other two components; however, as shown in Fig. 18, their effect on the resulting forecasting performance can be relatively significant (NRMSE mean ranging from 44.31% to 57.65% for D2, and from 24.46% to 32.47% for D1). The only exception is the D1 component for Tatranské Zruby, where the results for seven-day-ahead forecasting could not be used, and this component was considered to be effectively unpredictable for this

forecasting horizon. This is in accordance with previous results (using (S)ARMA models), where temperature regression could better explain the data variation than in the remaining buildings, leaving the random component in the residuals after temperature regression more significant. It is important to note that according to Table 4, only one regressor was required to obtain highly accurate forecasts of the first two wavelet components (A5 and D5) in each building except for D5 in Krompachy, with a slightly larger number for the remaining components. As shown in Table 4, the lowest number of regressors for each component (except for D1 in Tatranské Zruby) was A5(1), D5(1), D4(2), D3(6), and D4(7) for Tatranské Zruby, A5(1), D5(2), D4(2), D3(7), D2(10), and D1(13) for Krompachy, and A5(1), D5(1), D4(4), D3(6), D2(8), and D1(8) for Sečovce.

Location		**Tatranské Zruby**			**Krompachy**			**Sečovce**		
Comp.	*Value*	NRMSE [%]	MAE [m^3]	RS [-]	NRMSE [%]	MAE [m^3]	RS [-]	NRMSE [%]	MAE [m^3]	RS [-]
A5	**Mean**	99.44	0.0611	2.6	99.66	0.0675	2.4	99.73	0.0094	2.2
	Max/ Min	99.47/ 99.32	0.0747/ 0.0548	4/1	99.67/ 99.64	0.0716/ 0.0622	4/1	99.86/ 99.70	0.0107/ 0.0048	6/1
	Std	0.0431	0.0055	1.1738	0.0119	0.0027	0.966	0.0485	0.0017	1.6865
D5	**Mean**	97.82	0.2649	3.6	98.14	0.2889	3.7	98.17	0.0738	3.4
	Max/ Min	98.05/ 97.53	0.2964/ 0.2393	7/1	98.31/ 97.99	0.3131/ 0.2610	6/2	98.30/ 98.06	0.0827/ 0.0682	6/1
	Std	0.1488	0.0175	1.6465	0.121	0.0189	1.42	0.0780	0.0044	1.7127
D4	**Mean**	92.00	1.1841	6.3	86.62	1.5646	6.6	94.45	0.2740	7.1
	Max/ Min	92.39/ 90.09	1.4120/ 1.1258	9/2	87.38/ 85.68	1.6862/ 1.4445	10/2	94.87/ 93.87	0.3069/ 0.2577	10/4
	Std	0.6758	0.0825	1.9465	0.537	0.0866	2.59	0.2818	0.0148	2.1318
D3	**Mean**	92.58	1.3113	8.5	92.94	1.4179	10.4	90.31	0.4367	8.8
	Max/ Min	93.18/ 91.95	1.4305/ 1.2215	12/6	93.25/ 92.60	1.4682/ 1.3518	12/7	91.16/ 89.65	0.4765/ 0.3964	11/6
	Std	0.3508	0.0679	1.9579	0.199	0.0362	1.58	0.3852	0.0210	1.6193
D2	**Mean**	48.20	13.7670	14.7	44.31	21.4591	14.6	57.65	4.5229	16.3
	Max/ Min	50.18/ 46.05	14.4400/ 12.7722	19/7	46.41/ 41.41	22.3851/ 20.5205	19/10	59.91/ 55.76	4.7724/ 4.2493	21/8
	Std	1.2664	0.4408	3.8312	1.38	0.5630	3.20	1.1948	0.1395	3.4335
D1	**Mean**	-6.38	27.4850	23.9	24.46	23.6646	22.3	32.47	7.9793	19.7
	Max/ Min	2.12/ -11.95	28.8193/ 26.1235	33/15	28.20/ 20.37	25.0093/ 22.1430	35/13	35.12/ 29.73	8.3311/ 7.7386	26/8
	Std	4.1841	0.8264	5.4047	2.71	0.9928	6.52	1.7584	0.1907	5.6774

Table 4. Results of 10 runs of GA for regressor optimization of each wavelet decomposition component.

Based on these results, the final tests of the GA-optimized regWANN model were carried out using the models with the smallest number of regressors for each component and the best NRMSE values on the test data. The number of days differed for each dataset, but in this evaluation, the longer test datasets (Krompachy and Sečovce) were shortened to match the size of Tatranské Zruby, so that they could be directly compared. Thus, all datasets contained 76 days in total: from October 1, 2018, to December 15, 2018. The performance of selected models on the week-ahead forecasting of temperature regression residuals can be assessed using Fig. 18. The blue lines indicate the residuals of the test data obtained after performing temperature regression using the models from Table 2. Each column shows the successive addition of next-level detail forecasts, confirming their importance in improving the final forecasting accuracy. As mentioned previously, the D1 component for the

Tatranské Zruby series could not be reasonably forecasted for a given forecasting horizon, and thus this component was not used in the final model. The results of the temperature regression residual forecasting using the full models for Krompachy and Sečovce shown in Fig. 18 are also confirmed by the model fit values for the test data, which were 42.74% and 51.29%, respectively, compared with 19.67% for Tatranské Zruby.

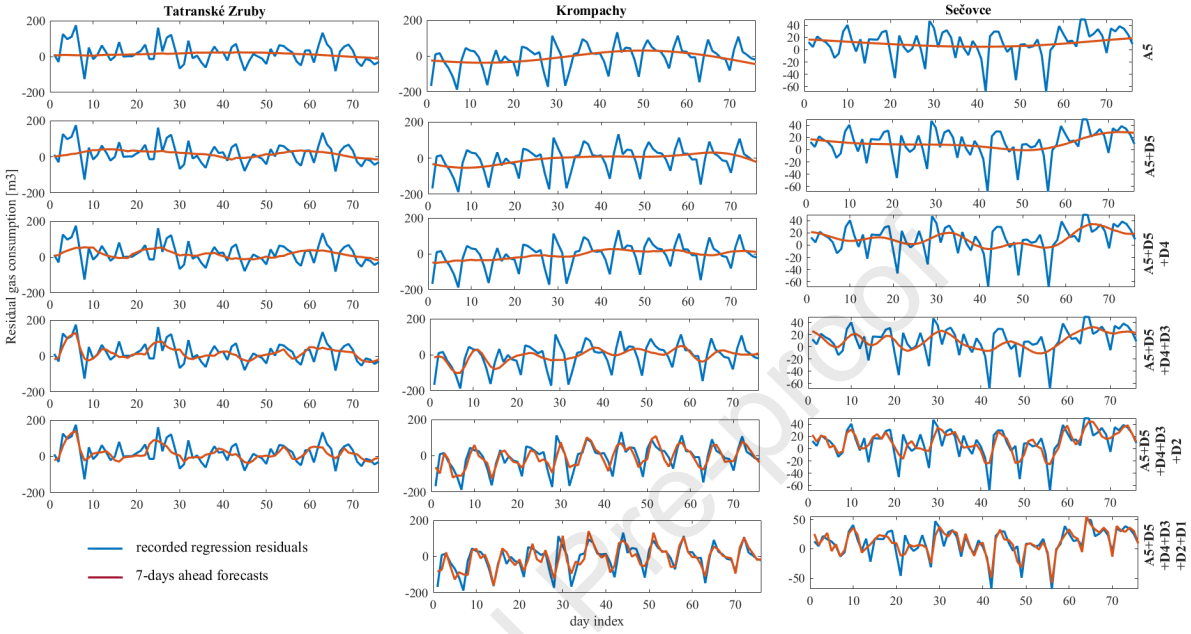


Fig. 18. Seven-day-ahead forecasting performance of GA-optimized regWANN models on regression residuals with the final reconstruction using all WT components (except for D1 for Tatranské Zruby residuals).

To test the performance of the GA-optimized regWANN model, the same 50-day dataset as in the case of the reg(S)ARMA models was used. The results of this test are shown in Fig. 19, where seven-day-ahead forecasting is again compared with the forecasts obtained using a simple regression model from Table 2. The regWANN model enabled a reduction of the MAE during the 50-day test data period to 37.44 m^3 (compared with 48.38 m^3 in the reg(S)ARMA approach) in Tatranské Zruby, that is, a 22.6% reduction. For the buildings in Krompachy and Sečovce, these values were reduced to 30.98 m^3 and 7.30 m^3 (compared with 37.65 m^3 and 16.97 m^3, respectively, for reg(S)ARMA), that is, a reduction by 17.7% and 57%, respectively. Even though Krompachy exhibited the smallest reduction, its NRMSE for the 50-day period remained better than that of Tatranské Zruby (71.0% compared with 61.7%). Likewise, as shown in the lower part of Fig. 19, the ranges of the maximum errors for any given day were significantly reduced, with only one of the errors exceeding +/− 100 m^3 for both Tatranské Zruby and Krompachy, and only three exceeding +/− 20 m^3 for Sečovce.

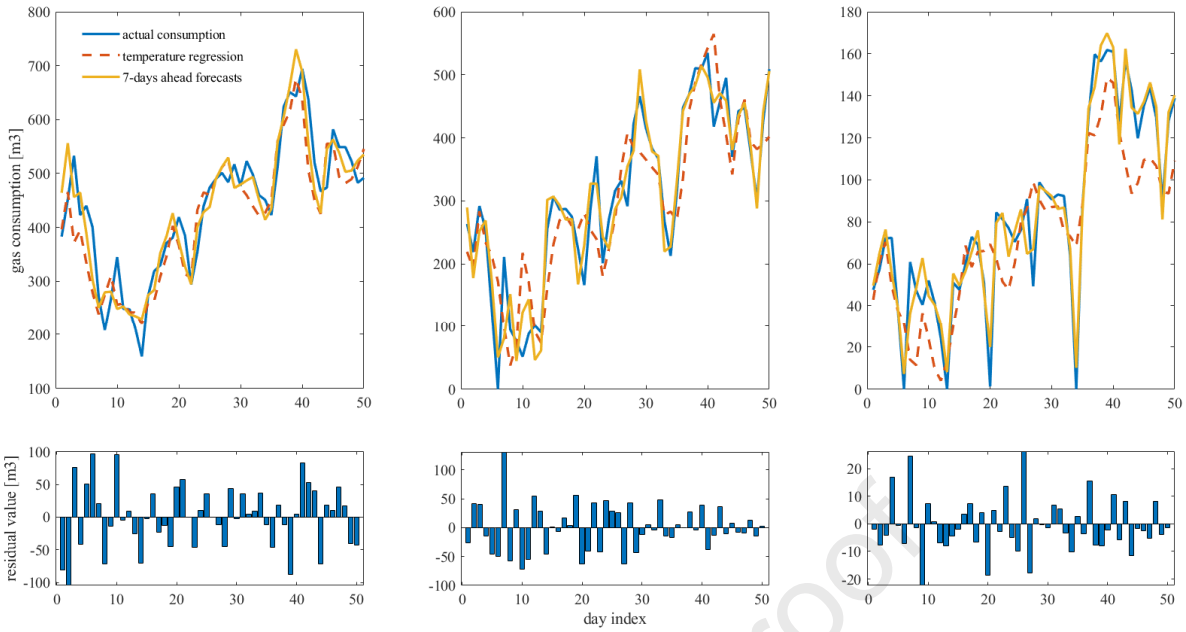


Fig. 19. Seven-day-ahead forecast of daily gas consumption using temperature regression model and WANN error models tested on 50-day test dataset and recorded temperatures.

To test the *ex ante* forecasting performance of regWANN models for a seven-day forecasting horizon, the same pre-sampling dataset as that shown in Fig. 15 was used. The results of week-ahead forecasting using regWANN models are shown in Fig. 20. It should be noted that the curves for actual consumption (blue line) and temperature regression (brown line) are the same as in Fig. 16. A striking difference is the improved forecasting accuracy of the regWANN models, including their ability to compensate for the incorrectly forecasted sudden drop in temperature. By observing the error bar graphs in Fig. 20, it is evident that the maximum absolute errors were reduced significantly to a maximum of approximately 100 m^3 in the case of Krompachy, and a maximum of approximately 50 and 40 m^3 in the case of Tatranské Zruby and Sečovce, respectively. The mean absolute errors for the entire period (Section III) were reduced to the following values: 30 m^3 (Tatranské Zruby), 69.35 m^3 (Krompachy), and 19.01 m^3 (Sečovce), that is, a reduction by 52.5%, 27%, and 43.6%, respectively.

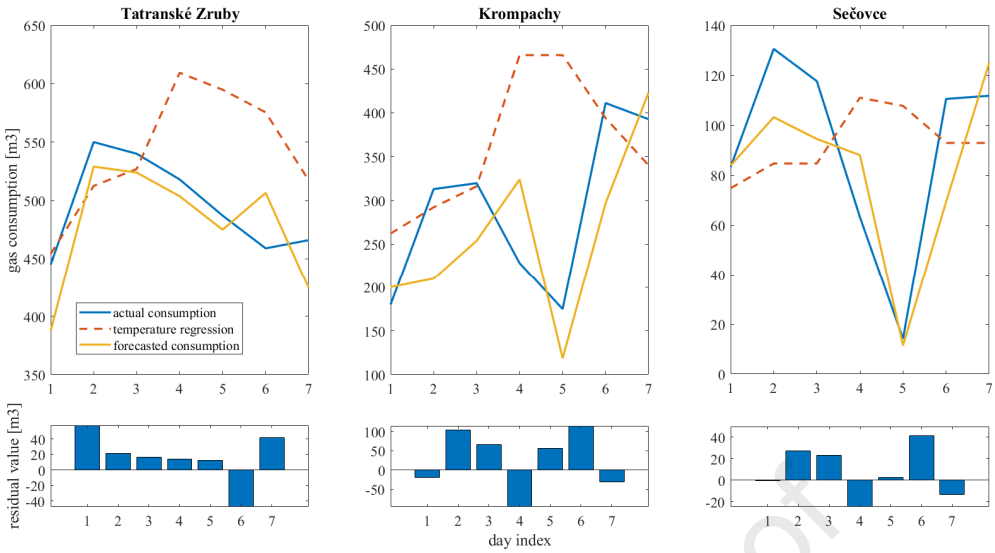


Fig. 20. Seven-day-ahead forecast of daily gas consumption using temperature regression model and WANN error models tested on 7-day test dataset and forecasted temperatures.

4.4 Summary of Results

The final results obtained for week-ahead forecasting of daily gas consumption can be evaluated and compared using both Table 5 and Fig. 21. This summary is divided into ex post and ex ante forecasting scenarios, corresponding to the previously introduced 50-day (ex post) and 7-day (ex ante) datasets. The 50-day test dataset started on October 23, 2018, and ended on December 11, 2018, whereas the seven-day test dataset started on March 4, 2020, and ended on March 10, 2020. It is important to recall that in the ex post forecasting scenario, recorded (rather than forecasted) temperatures were used, thus representing an ideal case from the perspective of temperature regression. In addition to the absolute values of MAE and NRMSE for each scenario, Table 5 shows the relative improvement of forecasting accuracy when the regWANN model was used. In three cases, this indicator was not used: The NRMSE values for (S)ARMA models in the ex ante scenario were negative (owing to the significant errors under these conditions) and could not be used. Furthermore, we did not use this indicator for minimum error comparison because low values of minimum error could, in general, be achieved by less accurate forecasting models at certain points as well. By contrast, the maximum error was more relevant, as it indicated the worst point forecast under the given conditions. In terms of MAE, the use of regWANN models resulted in improvements from 17.7% to 57%, which can be considered significant. Even though the MAE values indicate the differences between actual and forecasted values in meaningful units (cubic meters), the NRMSE values represent the closeness of the entire forecasted period to the recorded. Therefore, the NRMSE values in the ex ante forecasting scenario are incomparably lower than in ex post forecasts because the errors in

temperature forecasts had profound effects on the resulting forecasting accuracy. In this regard, the value of 38% in ex ante forecasts for Sečovce should be considered satisfactory. In addition, by using regWANN models, it was possible to reduce the maximum forecasting error for each of the buildings in the ex ante scenario by approximately 50%. This significant difference was primarily due to the aforementioned incorrectly forecasted sudden drop in the outdoor temperature, which was relatively successfully compensated for by the regWANN models.

It is also interesting to see the structure (inputs) of the final regWANN models after using the GA optimization of the regressor vector. As mentioned previously, the D1 component of Tatranské Zruby was not used because of its low predictability; hence, no regressors are shown. Even though the optimality of the obtained solution cannot be ensured when algorithms such as GA are used, the results shown in Table 5 represent the best balance between the total number of regressors and error in the test data.

Model	**Building**	**MAE** *ex post*	**NRMSE** *ex post*	**MAE** *ex ante*	**NRMSE** *ex ante*	**Max error** *ex post*	**Min error** *ex post*	**Max error** *ex ante*	**Min error** *ex ante*
(S)ARMA	T. Zruby	48.38 m^3	52.07 %	63.21 m^3		157.00m^3	2.21 m^3	118.38m^3	6.27 m^3
	Krompachy	37.65 m^3	62.22 %	94.90 m^3		177.63m^3	0.17 m^3	222.37m^3	14.57 m^3
	Sečovce	16.97 m^3	53.23 %	33.70 m^3		56.68 m^3	0.83 m^3	76.34 m^3	5.34 m^3
WANN	T. Zruby	37.44 m^3	61.71 %	30.00 m^3	9.96 %	104.27m^3	0.01 m^3	57.54 m^3	12.03 m^3
	Krompachy	30.98 m^3	71.01 %	69.35 m^3	13.57 %	130.05m^3	0.18 m^3	114.31m^3	19.93 m^3
	Sečovce	7.30 m^3	78.59 %	19.01 m^3	37.96 %	26.10 m^3	0.38 m^3	41.19 m^3	0.60 m^3
Improvement	T. Zruby	22.6%	9.64 %	52.5 %		33.6 %		51.4 %	
	Krompachy	17.7 %	8.79 %	27 %		26.8 %		48.6 %	
	Sečovce	57 %	25.36 %	43.6 %		54.0 %		46.0 %	

regWANN regressors	**A5**	**D5**	**D4**	**D3**	**D2**	**D1**
Tat.Zruby	*y*(*k*-5)	*y*(*k*-20)	*y*(*k*-26), *y*(*k*-38)	*y*(*k*-8), *y*(*k*-21), *y*(*k*-25), *y*(*k*-41), *y*(*k*-42), *y*(*k*-53)	*y*(*k*-8), *y*(*k*-13), *y*(*k*-19), *y*(*k*-25), *y*(*k*-43), *y*(*k*-56), *y*(*k*-63)	
Krompachy	*y*(*k*-18)	*y*(*k*-9), *y*(*k*-12)	*y*(*k*-28), *y*(*k*-44)	*y*(*k*-5), *y*(*k*-7), *y*(*k*-10), *y*(*k*-35), *y*(*k*-38), *y*(*k*-45), *y*(*k*-54), *y*(*k*-59)	*y*(*k*-9), *y*(*k*-13), *y*(*k*-23), *y*(*k*-25), *y*(*k*-37), *y*(*k*-39), *y*(*k*-51), *y*(*k*-63), *y*(*k*-66), *y*(*k*-69)	*y*(*k*-13), *y*(*k*-21), *y*(*k*-22), *y*(*k*-24), *y*(*k*-25), *y*(*k*-28), *y*(*k*-33), *y*(*k*-37), *y*(*k*-48), *y*(*k*-51), *y*(*k*-56), *y*(*k*-68), *y*(*k*-73)
Sečovce	*y*(*k*-14)	*y*(*k*-19)	*y*(*k*-16), *y*(*k*-21), *y*(*k*-30), *y*(*k*-34)	*y*(*k*-13), *y*(*k*-45), *y*(*k*-46), *y*(*k*-52), *y*(*k*-38), *y*(*k*-45), *y*(*k*-53), *y*(*k*-57)	*y*(*k*-1), *y*(*k*-3), *y*(*k*-11), *y*(*k*-24), *y*(*k*-28), *y*(*k*-36), *y*(*k*-62), *y*(*k*-66)	*y*(*k*-3), *y*(*k*-4), *y*(*k*-18), *y*(*k*-22), *y*(*k*-24), *y*(*k*-33), *y*(*k*-41), *y*(*k*-47)

Table 5. Summary of the results obtained for week-ahead gas consumption forecasting using reg(S)ARMA and regWANN models and regressors in the final regWANN models.

The performance of the reg(S)ARMA and regWANN models for week-ahead gas consumption forecasting can be evaluated using Fig. 21, which shows the absolute daily forecasting errors for both test datasets (50-day and 7-day). The values above the x-axis represent the errors for regWANN models, whereas the values below the x-axis were obtained using the reg(S)ARMA models. To facilitate the comparison, the red color indicates errors with larger magnitude, whereas the green color indicates smaller errors. A direct comparison of the daily week-

ahead forecasting performance of the models confirms the lower average error for the regWANN model as well as the lower maximum errors in both forecasting scenarios and for each building. The total number of lower daily errors was 38 (Tatranské Zruby), 27 (Krompachy), and 39 (Sečovce) for ex post forecasts, and 5 (Tatranské Zruby), 3 (Krompachy), and 5 (Sečovce) for ex ante forecasts. For both models, the higher values of forecasting errors were clearly due to sudden changes in weather conditions, which were not reflected in the weather forecasts (e.g., the situation around the end of October 2018). This was most pronounced for Krompachy, where the maximum error values were the highest for both model types and in both forecasting scenarios. In addition, in this case, the differences in the model forecasting performance was less significant (similar numbers of lower absolute daily errors and less remarkable improvement). However, the best performance of the regWANN model and the most striking improvement over the reg(S)ARMA model could be observed for Sečovce. As shown in Figs. 13 and 18, the modeling of errors after regression had a profound effect on the forecasting accuracy for this building (the difference in forecasted values between the simple temperature regression model and the model with modeled residuals). Naturally, the accuracy of weather forecasts plays an important role in the resulting forecasting performance of both types of models. This was particularly visible in the NRMSE values of regWANN models for ex post and ex ante forecasts, with the former being several times higher.

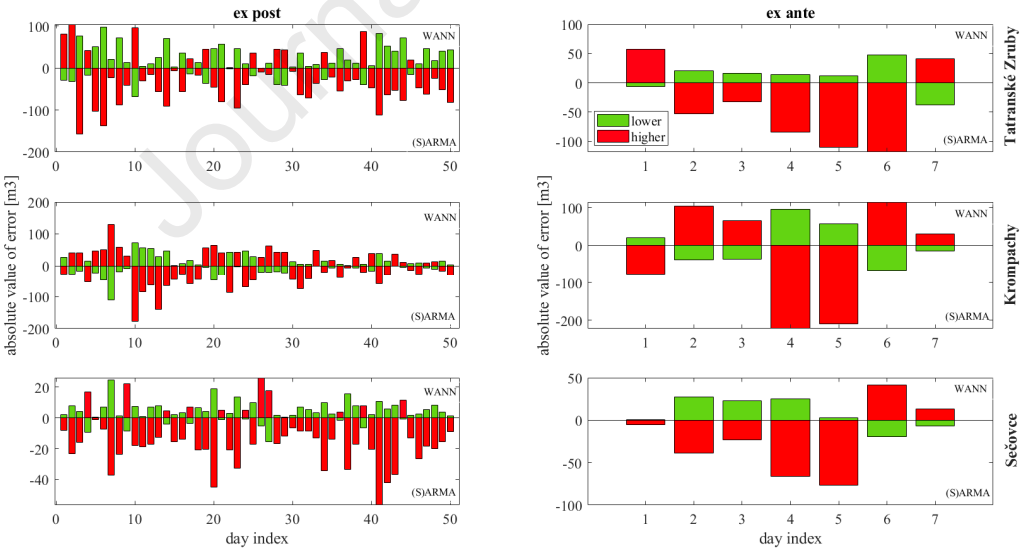


Fig. 21. Comparison of absolute values of errors of week-ahead daily forecasts in *ex post* and *ex ante* forecasting scenarios.

The significance of the obtained results with regard to real-world conditions can be related to effective energy management in various types of buildings. This effect is expected to be observed not only in terms of energy costs but also in heat production efficiency control. Specifically, the forecasting model developed here is intended to be used for the following purposes in a practical scenario:

1. With changing conditions that affect the gas market, the dynamic purchasing of gas offers greater flexibility for gas consumers. It is expected that the dynamically changing maximum (as well as the minimum) daily gas consumption will be contracted under dynamic purchasing for each day of the week also for small and medium-sized gas consumers in near future. In such a scenario, the accuracy of gas consumption forecasting will be of crucial importance and even small improvements of its value can have significant benefits. Even though the forecasting horizon may differ depending on actual conditions and requirements, it is reasonable to assume that the performance improves for shorter horizons. On the other hand, in the setting of daily gas consumption in buildings, the importance of temperature in its forecasting is paramount and therefore the week-ahead horizon can be considered an acceptable maximum for reliable temperature forecasts. This applies despite the fact that the GA-optimized regWANN model is shown to be capable of compensating for the inaccuracies in temperature forecasts.
2. In addition, the model is planned to be used for efficiency control of heat sources in different types of buildings. In this case, it is expected that the possible forecasting horizon may be shorter and the operation of a given heat source is evaluated using the results of daily gas consumption forecasting. Should the discrepancy between the forecasted and actual value exceed the defined threshold, the operation of a heat source may be subject to further investigation.

5 Conclusion

In this paper, a comparative study of week-ahead forecasting of daily gas consumption in three different types of buildings was presented using two types of models. To demonstrate the more general usability of the forecasting models, three buildings with different purposes (and gas consumption profiles) were selected: a hotel, an elementary school, and a healthcare building. As daily gas consumption is strongly correlated with the average outdoor temperature, the tested models used temperature regression with different modeling of the serially correlated residuals (ARMA/SARMA modeling vs. GA-optimized WANN modeling). In addition, both models were tested in two forecasting scenarios: *ex post* (where recorded temperatures were used) and *ex ante* (where week-ahead forecasted temperatures were used). Based on the results, the following conclusions can be drawn: I. Regression-based forecasting with separate modeling for the time series of the residuals is preferable to the forecasting of the original daily gas consumption time series. In both cases, the residual modeling significantly reduced the forecasting error compared with a simple temperature regression. II. The temporal characterization

of a given gas consumption profile may have a noticeable effect on forecasting accuracy. III. The use of the GA-optimized regWANN approach for daily gas-consumption forecasting may improve forecasting accuracy, even under conditions of more significant temperature forecasting errors.

It was demonstrated that the WT could be used to isolate components with different cycle periods, which are present in the original time series, and this improves forecasting accuracy. As more flexible models were used, it was important to include an algorithm to search for a satisfactory (not necessarily optimal) selection of regressors in the model. Moreover, the resulting regWANN is naturally more complex than reg(S)ARMA models. However, we do not expect this to be an issue in most cases where sufficient (but not prohibitively large) computational power is large.

There are several aspects that could be addressed in future research regarding the possible use of regWANN models to forecast daily gas consumption in an online scenario. As regWANN is a temperature-regression-based model, it would be useful to further examine the possibilities for improving the accuracy of average daily temperature forecasts. Also, to achieve even better parsimony of the models, a more rigorous approach to the selection of wavelet decomposition levels could be researched. This can be combined with the testing of different search algorithms for finding further improvements in forecasting accuracy.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0602 and by the Ministry of Industry and Trade of the Czech Republic project No. FV20419.

References

[1] H. Zhao, F. Magoules, A review on the prediction of building energy consumption, Renew. Sust. Energ. Rev. 16 (2012) 3586-3592. https://doi.org/10.1016/j.rser.2012.02.049

[2] T. Ahmad, H. Chen, Y. Guo, J, Wang, A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. Energy Build. 165 (2018) 301-320. https://doi.org/10.1016/j.enbuild.2018.01.017

[3] L. Suganthi, A. A. Samuel, Energy models for demand forecasting – A review. Renew. Sust. Energ. Rev. 16 (2012) 1223-1240. https://doi.org/10.1016/j.rser.2011.08.014

[4] S. Idowu, S. Saguna, C. Ahlund, O. Schelén, Applied machine learning: Forecasting heat load in district heating system, Energy Build. 133 (2016) 478-488. https://doi.org/10.1016/j.enbuild.2016.09.068

[5] C. Deb, L. S. Eang, J. Yang, M. Santamouris, Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. Energy Build. 121 (2016) 284-297. https://doi.org/10.1016/j.enbuild.2015.12.050

[6] F. Taspinar, N. Celebi, N. Tutkun, Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods. Energy Build. 56 (2013) 23-31. https://doi.org/10.1016/j.enbuild.2012.10.023

[7] H. R. Khosravani et al., A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building. Energies 9 (2016). https://doi.org/10.3390/en9010057

[8] R. Ž. Jovanović, A. A. Sretenović, B. D. Živković, Ensemble of various neural networks for prediction of heating energy consumption, Energy Build. 94 (2015) 189-199. https://doi.org/10.1016/j.enbuild.2015.02.052

[9] K. Zhang, R. Gencay, M.E. Yazgan, Application of wavelet decomposition in time-series forecasting, Econ. Lett. 158 (2017) 41-46. https://doi.org/10.1016/j.econlet.2017.06.010

[10] M. Rana, I. Koprinska, Forecasting electricity load with advanced wavelet neural networks. Neurocomputing 182 (2016) 118-132. https://doi.org/10.1016/j.neucom.2015.12.004

[11] H. Zhang et al., Forecasting of PM_{10} time series using wavelet analysis and wavelet-ARMA model in Taiyuan, China. **J. Air. Waste Manag. Assoc.** 67 (2017) 776-788. https://doi.org/10.1080/10962247.2017.1292968

[12] A. Meng, J. Ge, H. Yin, S. Chen, Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm, Energy Convers. Manag. 114 (2016) 75-88. https://doi.org/10.1016/j.enconman.2016.02.013

[13] I. P. Panapakidis, A. S. Dagoumas, Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. Energy 118 (2017) 231-245. https://doi.org/10.1016/j.energy.2016.12.033

[14] R. Maheswaran, R. Khosa, Wavelet Volterra Coupled Models for forecasting of Nonlinear and Non-stationary time series. Neurocomputing 149 (2015) 1074-1084. https://doi.org/10.1016/j.neucom.2014.07.027

[15] H. Su et al, A hybrid hourly natural gas demand forecasting method based on integration of wavelet transform and enhanced Deep-RNN model. Energy 178 (2019) 585-597. https://doi.org/10.1016/j.energy.2019.04.167

[16] T. Liu et al., Study on deep reinforcement learning techniques for building energy consumption forecasting. Energy Build. 208 (2020) 109675. https://doi.org/10.1016/j.enbuild.2019.109675

[17] G. Song, Q. Dai, A novel double deep ELMs ensemble system for time series forecasting. Knowl.-Based Syst. 134 (2017) 31-49. https://doi.org/10.1016/j.knosys.2017.07.014

[18] G. Chitalia, M. Pipattanasomporn, V. Garg, S. Rahman, Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks. Appl. Energy 278 (2020) 115410. https://doi.org/10.1016/j.apenergy.2020.115410

[19] A. Sagheer, M. Kotb, Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing 325 (2019) 203-213. https://doi.org/10.1016/j.neucom.2018.09.082

[20] A. Vaghefi, M.A. Jafari, E. Bisse, Y. Lu, J. Brouwer, Modeling and forecasting of cooling and electricity load demand, Appl. Energy 136 (2014) 186-196. https://doi.org/10.1016/j.apenergy.2014.09.004

[21] M. Akpinar, N. Yumusak, Year ahead demand forecast of city natural gas using seasonal time series methods, Energies 9 (9) (2016). https://doi.org/10.3390/en9090727

[22] G. Dudek, Pattern similarity-based methods for short-term load forecasting – Part 1: Principles, Appl. *Soft* Comput. 37 (2015) 277-287. https://doi.org/10.1016/j.asoc.2015.08.040

[23] G. Dudek, Pattern similarity-based methods for short-term load forecasting – Part 2: Models, Appl. *Soft* Comput. 36 (2015) 422-441. https://doi.org/10.1016/j.asoc.2015.07.035

[24] H. Zhao, F. Magoules, Feature Selection for Predicting Building Energy Consumption Based on Statistical Learning Method. J. Algorithm. Comput. Technol. 6 (2012) 59-77. https://doi.org/10.1260/1748-3018.6.1.59

[25] E. Yukseltan, A. Yucekaya, A. H. Bilge, Hourly electricity demand forecasting using Fourier analysis with feedback. Energy Strategy Rev. 31 (2020) 100524. https://doi.org/10.1260/1748-3018.6.1.59

[26] P. Lusis, K. R. Khalilpour, L. Andrew, A. Liebman, Short-term residential load forecasting: Impact of calendar effects and forecast granularity. Appl. Energy 205 (2017) 654-669. https://doi.org/10.1016/j.apenergy.2017.07.114

[27] P. Bacher, H. Madsen, H.A. Nielsen, B. Perers, Short-term heat load forecasting for single family houses, Energy Build. 65 (2013) 101-112. https://doi.org/10.1016/j.enbuild.2013.04.022

[28] E. M. de Oliveira, F. L. C. Oliveira, Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods. Energy 144 (2018) 776-788. https://doi.org/10.1016/j.energy.2017.12.049

[29] M. del C. Bas, J. Ortiz, L. Ballesteros, S. Martorell, Evaluation of a multiple linear regression model and SARIMA model in forecasting ^{7}Be air concentrations. Chemosphere 177 (2017) 326-333. https://doi.org/10.1016/j.chemosphere.2017.03.029

[30] M. Bouzerdoum, A. Mellit, A. M. Pavan, A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. Sol. Energy 98 (2013) 226-235. https://doi.org/10.1016/j.solener.2013.10.002

[31] A. Hošovský, J. Piteľ, J. Mižáková, K. Židek, Introductory Analysis of Gas Consumption Time Series in Non-residential Buildings for Prediction Purposes Using Wavelet Decomposition. MM Sci. J. (2018). https://doi.org/10.17973/MMSJ.2018.12.201858

[32] J. Piteľ, J. Mižáková, A. Hošovský, Biomass combustion control and stabilization using low-cost sensors. Adv. Mech. Eng. 2013 (2013). https://doi.org/10.1155/2013/685157

[33] J. Mižáková et al., Using special filter with membership function in biomass combustion process control. Appl. Sci. 8 (2018). https://doi.org/10.3390/app8081279

[34] I. Čorný, Overview of Progressive Evaluation Methods for Monitoring of Heat Production and Distribution. Procedia Eng. 190 (2017) 619-626. https://doi.org/10.1016/j.proeng.2017.05.388

[35] Box, E.P.G., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. Time Series Analysis – Forecasting and Control, John Wiley & Sons, Hoboken, 2016. https://doi.org/10.1111/jtsa.12194

[36] W. Palma, Time Series Analysis, John Wiley & Sons, Hoboken, 2016

[37] Econometric Toolbox User's Guide, The MathWorks Inc., 2018b

[38] A. K. Tangirala, Principles of System Identification, CRC Press, Boca Raton, 2015

[39] S. Soltani, On the use of the wavelet decomposition for time series prediction. Neurocomputing 48 (2002) 267-277. https://doi.org/10.1016/S0925-2312(01)00648-8

[40] A. K. Alexandridis, A. D. Zapranis, Wavelet Neural Networks, John Wiley & Sons, Hoboken, 2014

[41] M. Stachura, Detecting Seasonality via Wavelet Methods. Studia Ekonomiczne 207 (2014) 223-232

[42] G. Chevillon, G.F. Hendry, Non-parametric direct multi-step estimation for forecasting economic processes. Int. J. Forecast. 21 (2005) 201-218. https://doi.org/10.1016/j.ijforecast.2004.08.004

[43] S.B. Taieb, R.J. Hyndman, Recursive and direct multi-step forecasting: the best of both worlds, Monash University, working paper, 2012

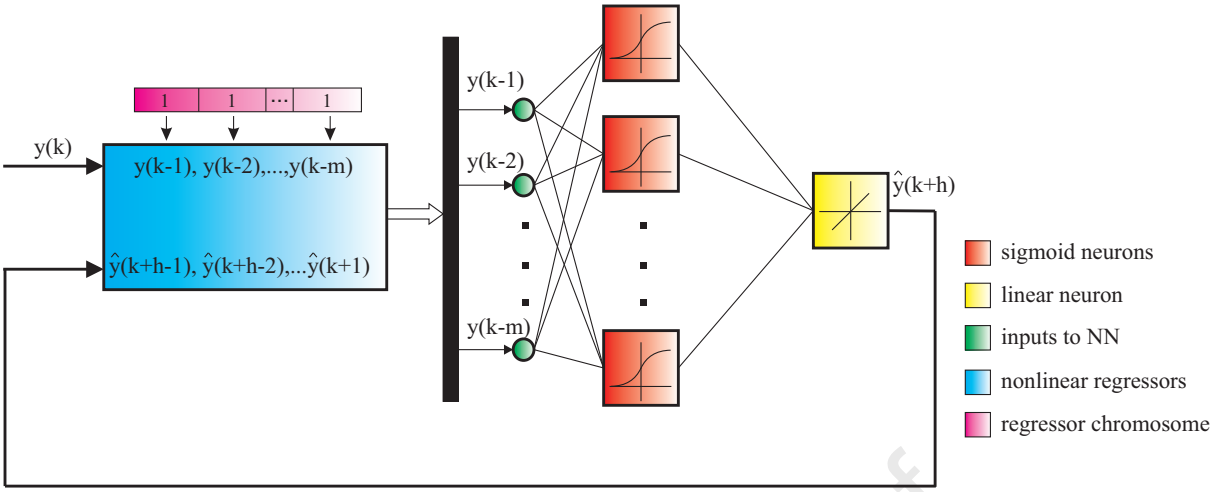
[44] S. Haykin, Neural Networks and Learning Machines, Prentice Hall, New York, 2009

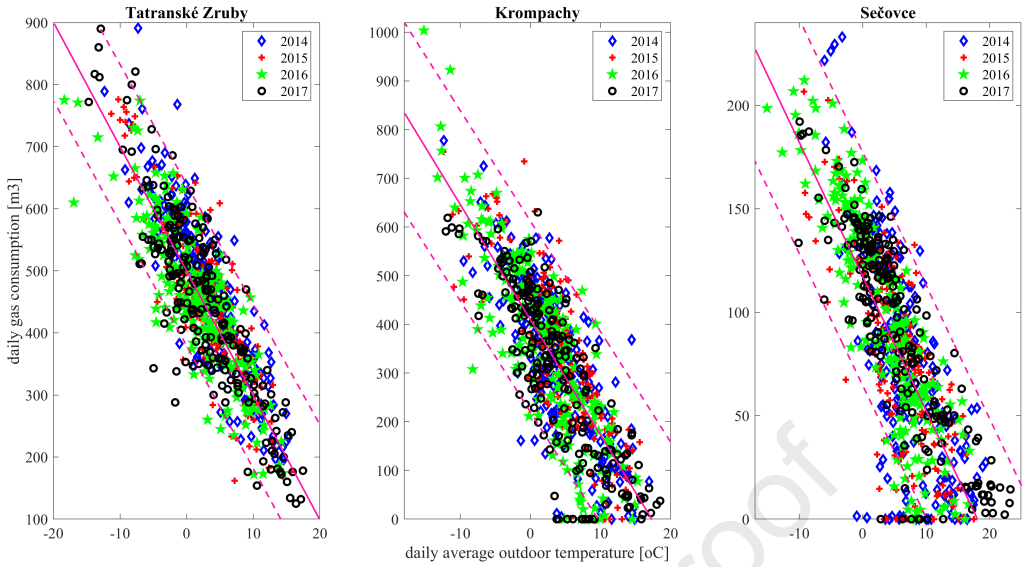
[45] M. H. Beale, M. T. Hagan, H. B. Demuth, Deep Learning Toolbox User's Guide, The MathWorks Inc. 2018b

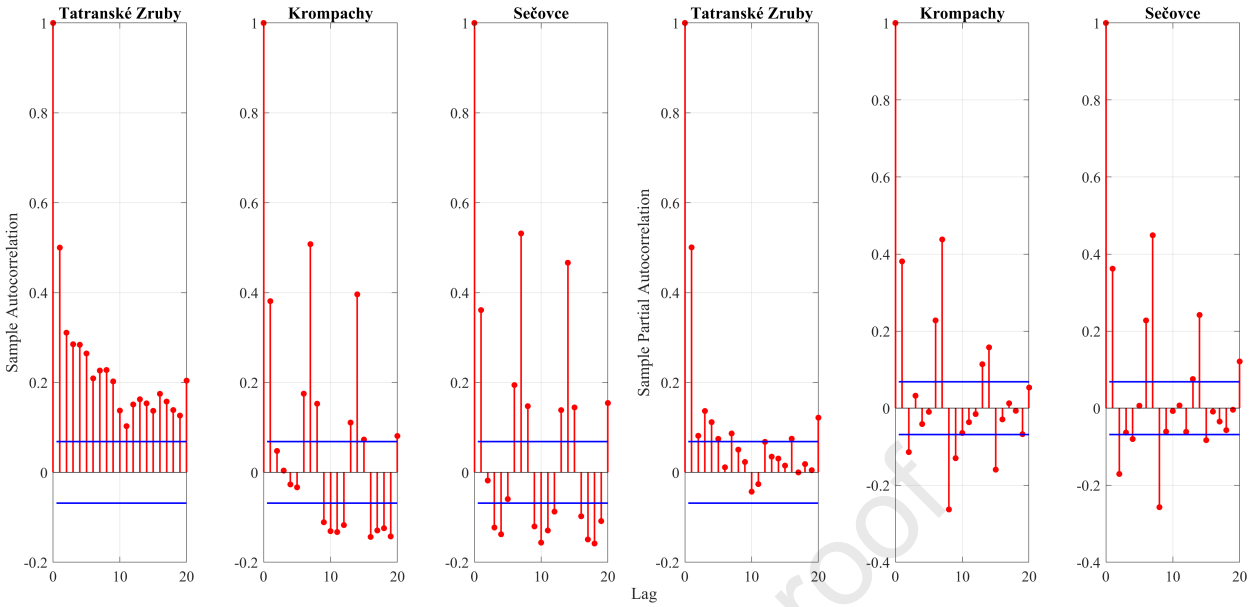
[46] J. Espinosa, J. Vandewalle, V. Wertz, Fuzzy Logic, Identification and Predictive Control, Springer, Berlin, 2005

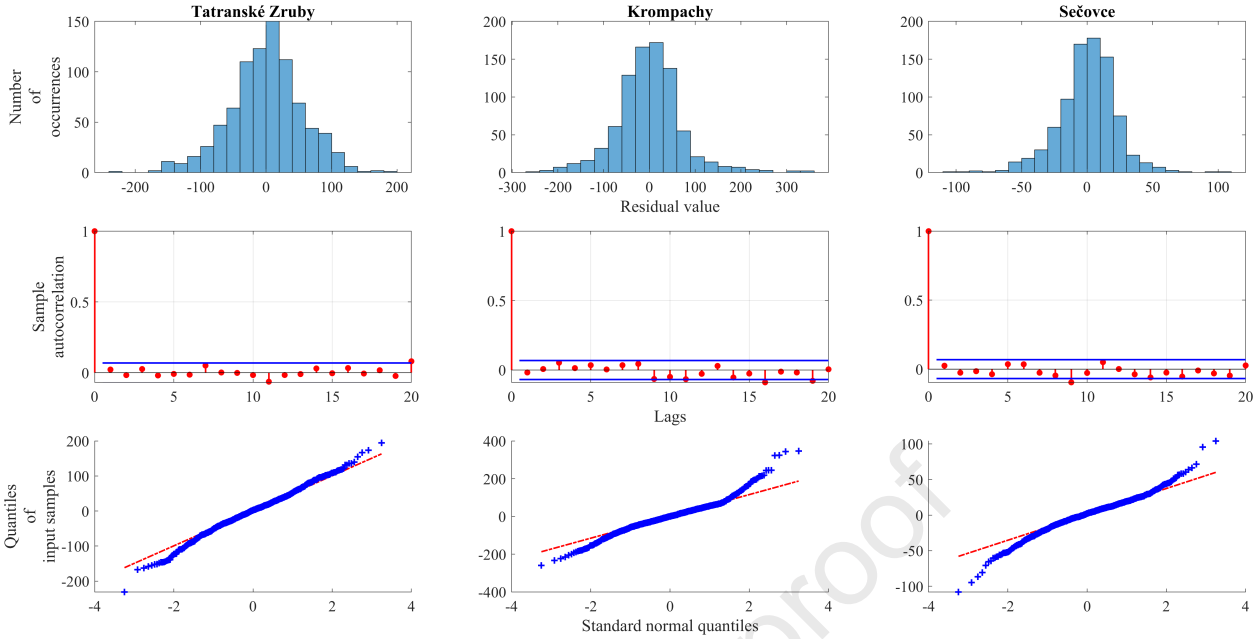
[47] K. Deep, K. P. Singh, M. L. Kansal, C. Mohan, A real coded genetic algorithm for solving integer and mixed integer optimization problems. Appl. Math. Comput. 212 (2009) 505-518. https://doi.org/10.1016/j.amc.2009.02.044

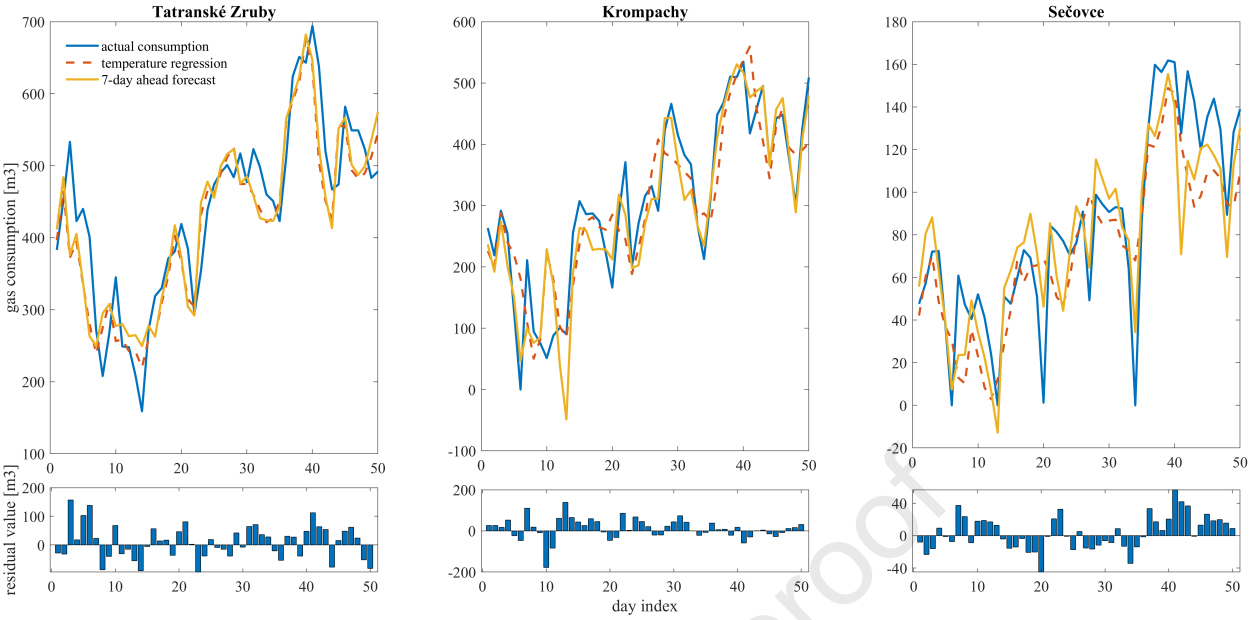
[48] Official webpage of Norwegian Meteorological Institute, www.yr.no, last access: August 31, 2020

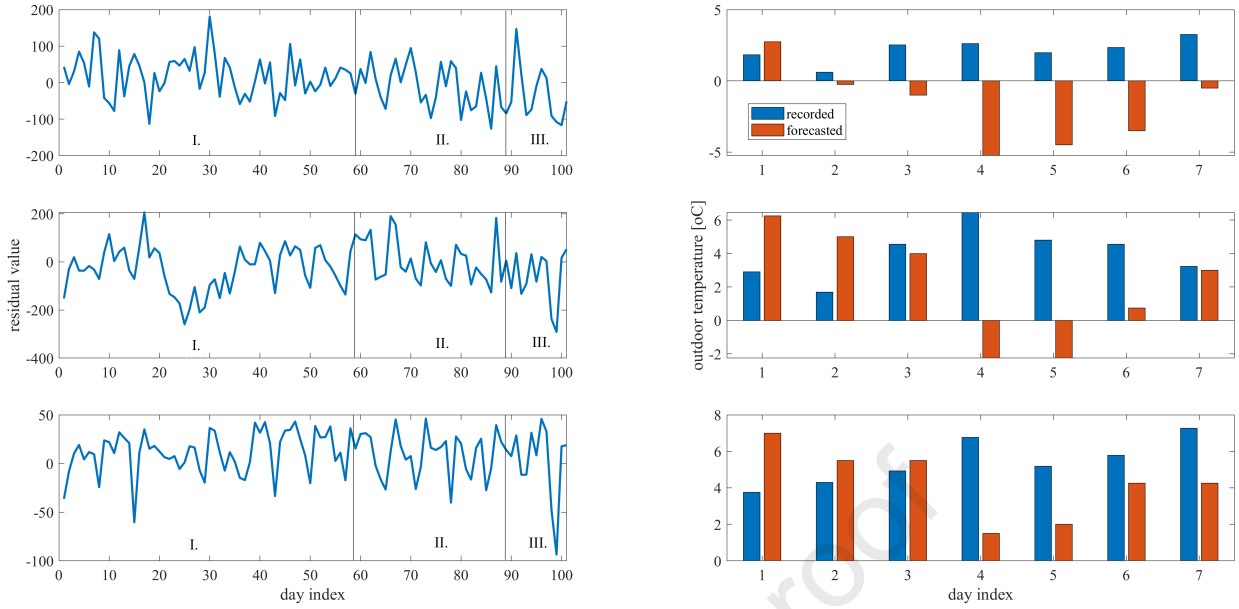


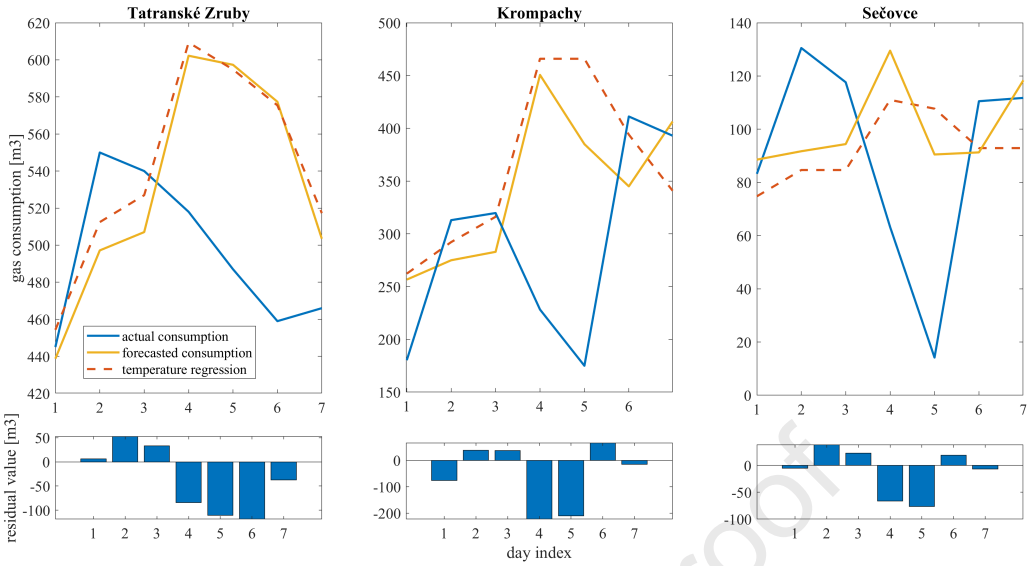


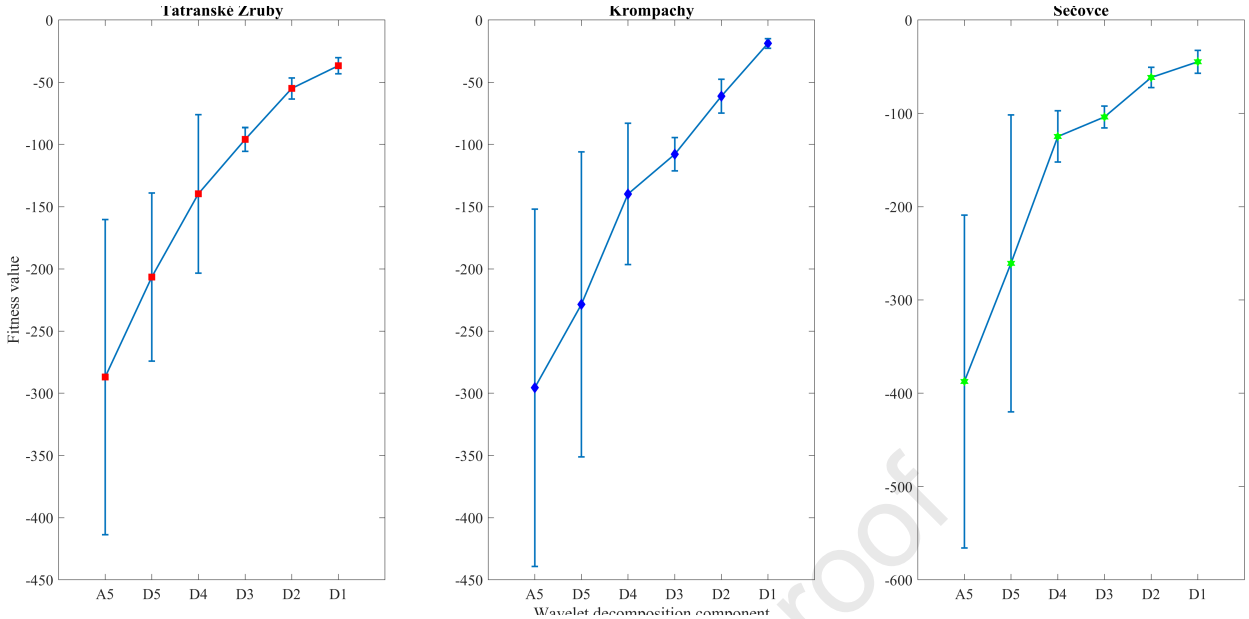


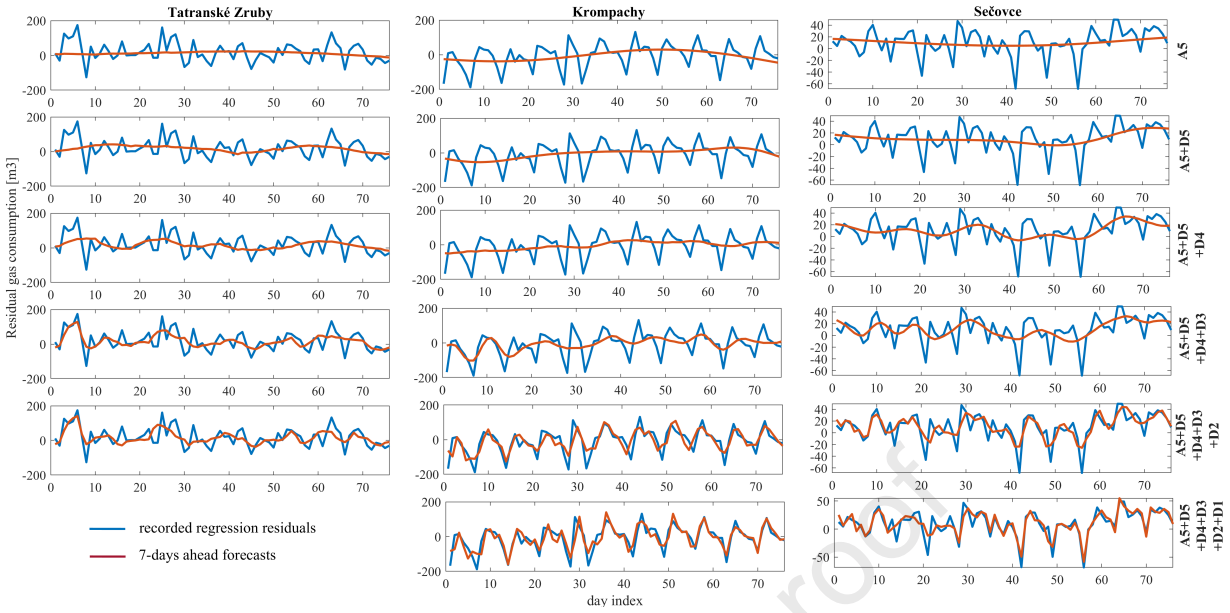


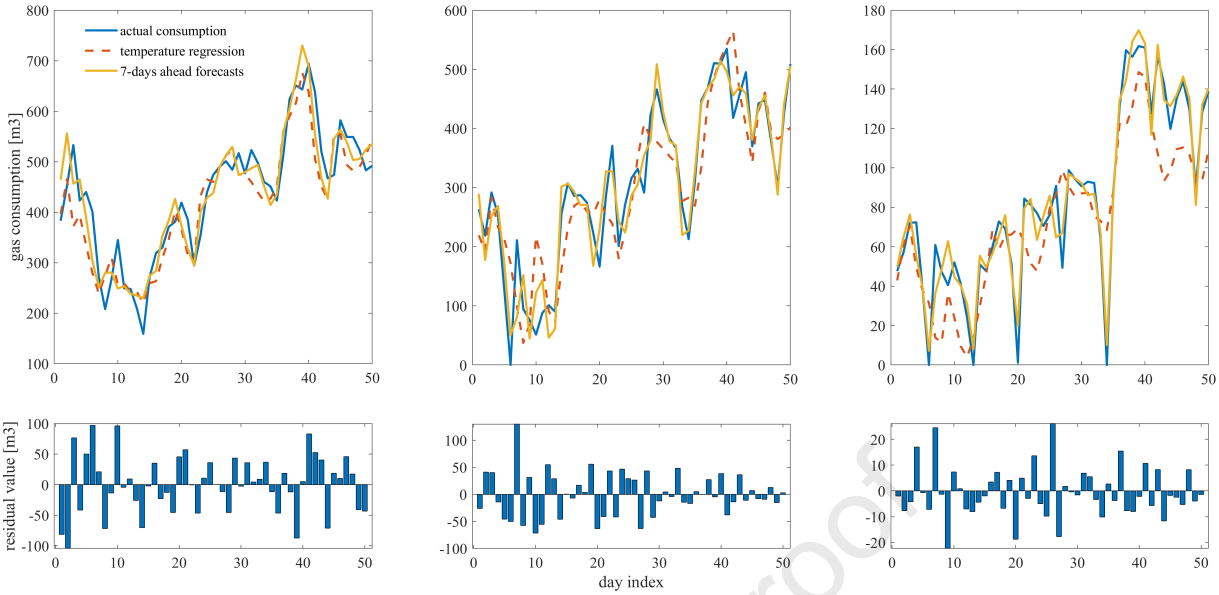


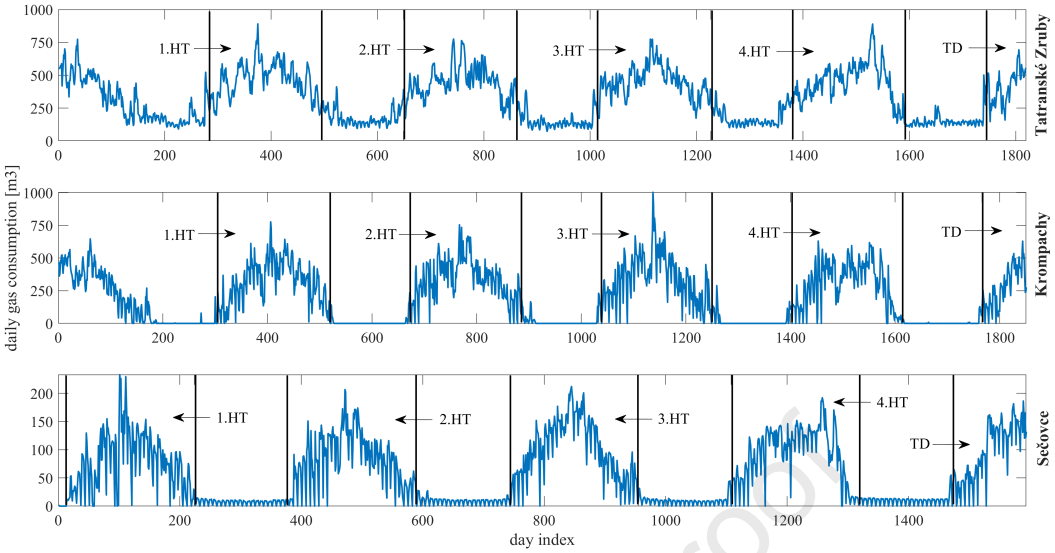


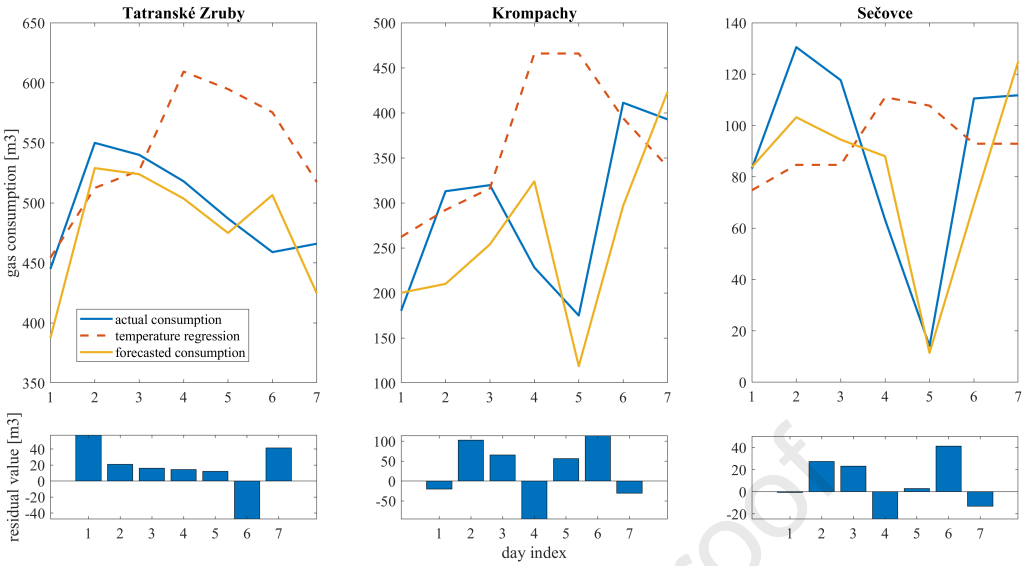


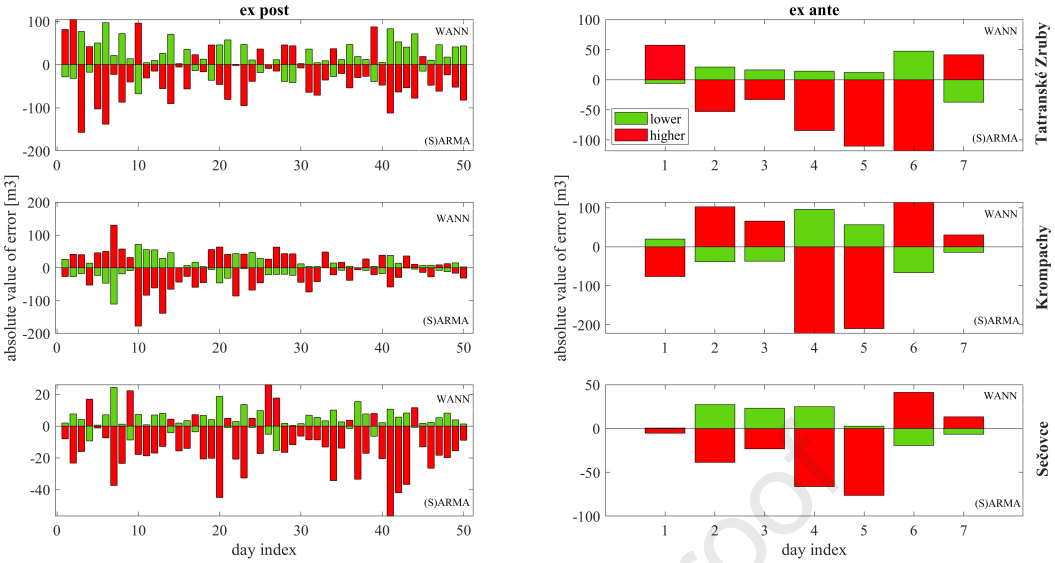


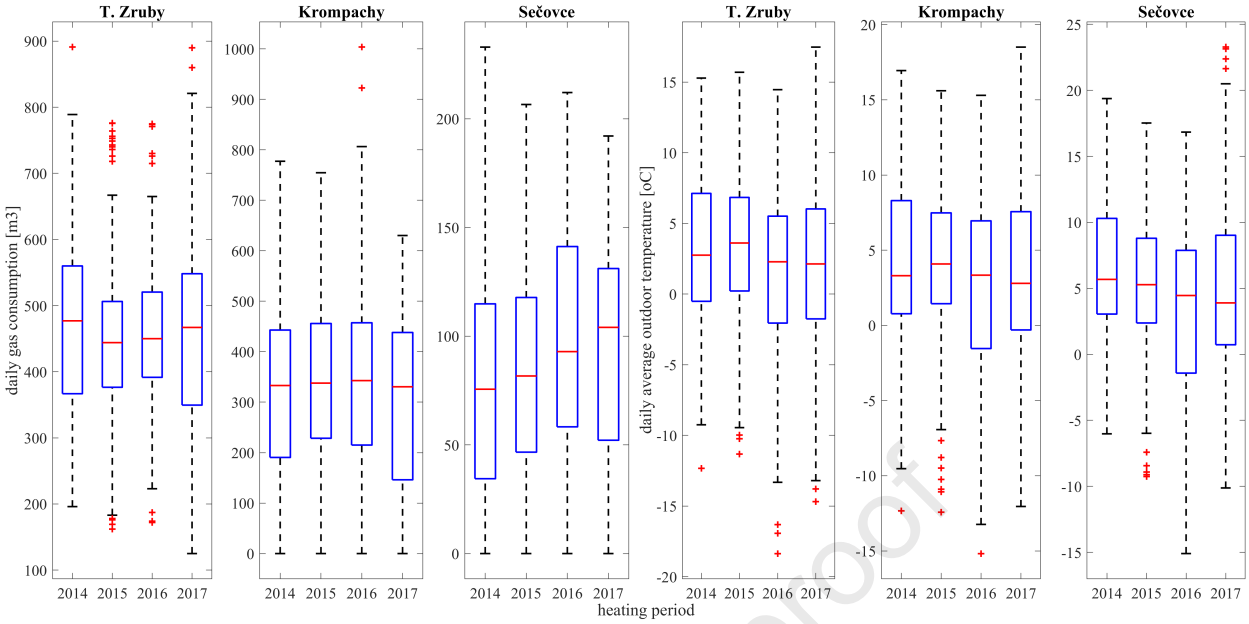


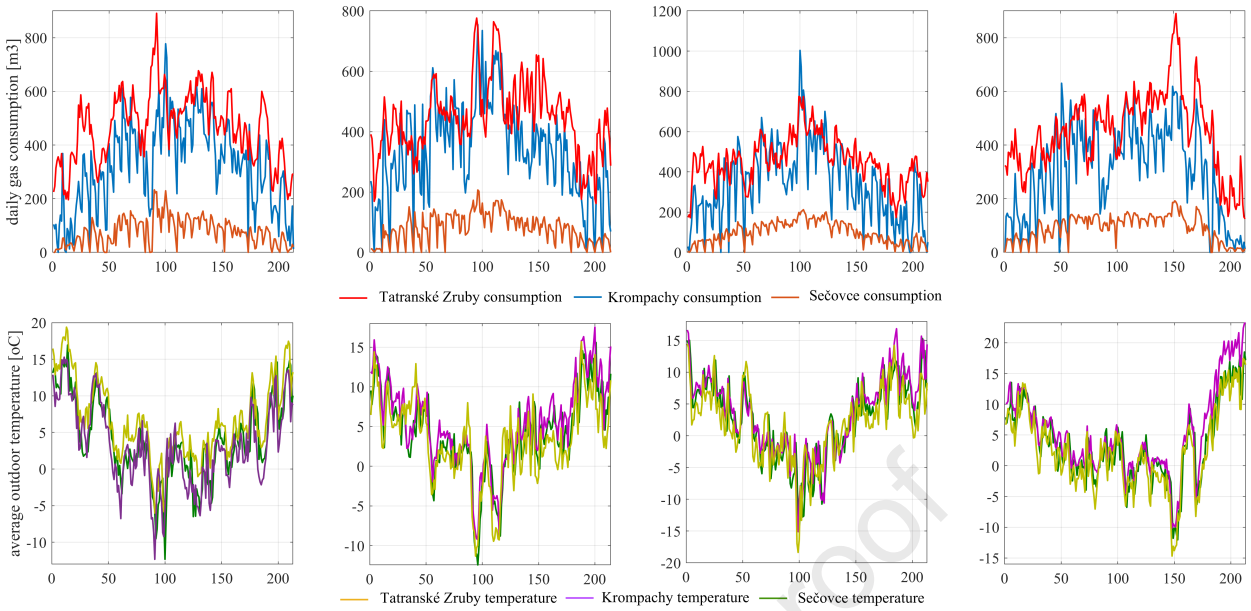


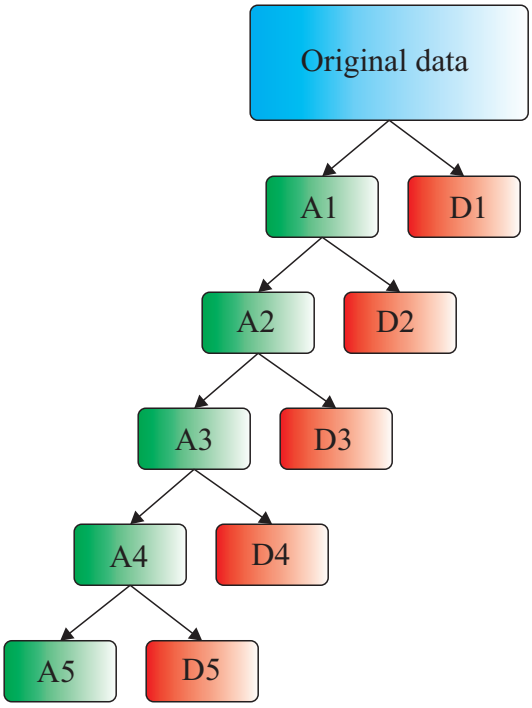


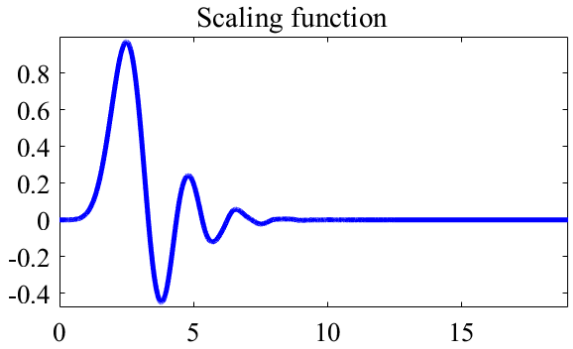


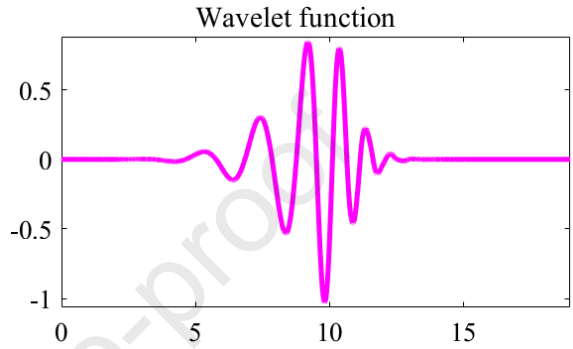


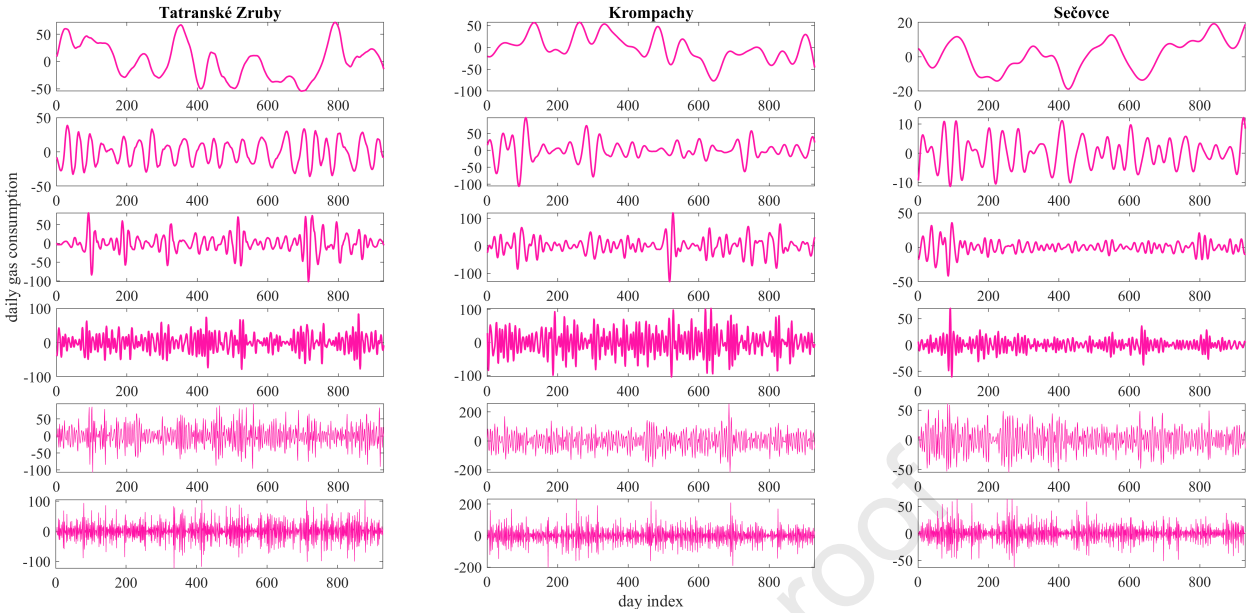


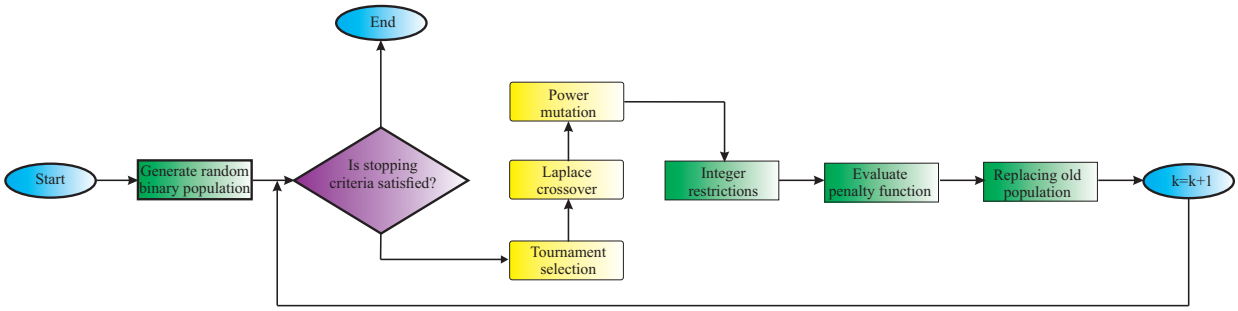




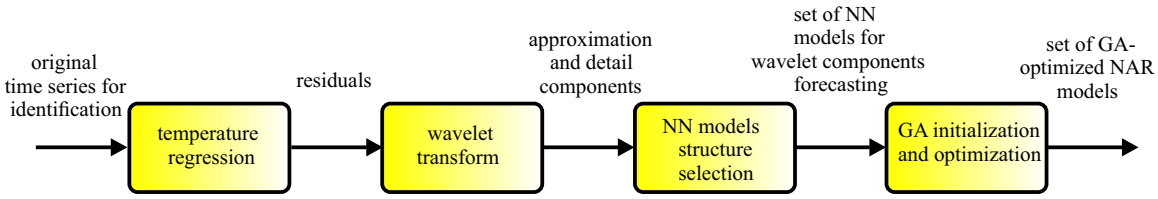




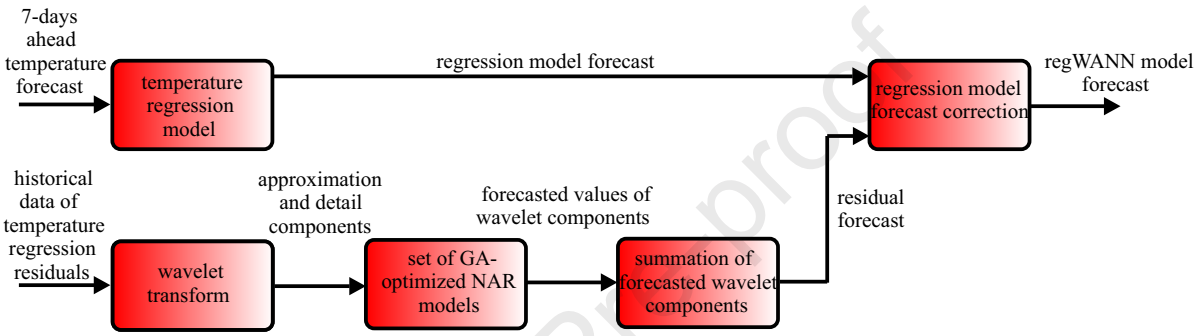


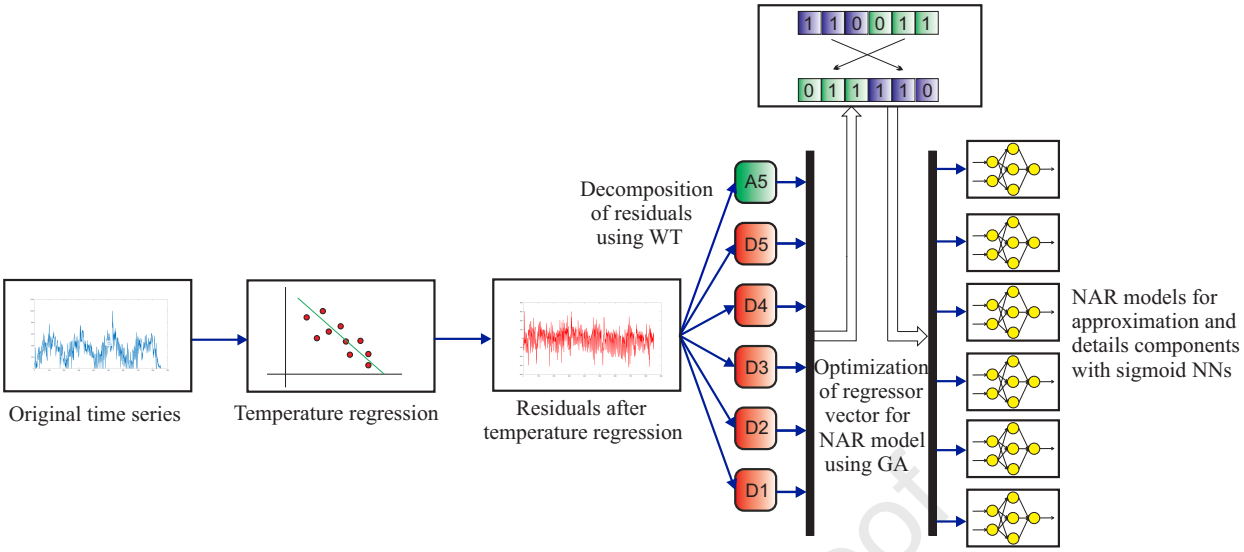


Identification



Forecasting





- Temperature regression using linear model offers reasonable gas consumption estimate
- Complex seasonality of gas consumption is handled effectively with db wavelets
- Binary GA is found instrumental in selecting near optimal regressors for NAR models
- Use of GA-optimized regWANN model compensates for temperature forecast inaccuracies

Fig.1 – two column

Fig.2 – one column

Fig.3 – one column

Fig.4 – one-and-half column

Fig.5 – one column

Fig.6 – two column

Fig.7 – two column

Fig.8 – one-and-half column

Fig.9 – one column

Fig.10 – one column

Fig.11 – one column

Fig.12 – one column

Fig.13 – one-and-half column

Fig.14 - one-and-half column

Fig.15 – one column

Fig.16 – one column

Fig.17 – one column

Fig.18 – two column

Fig.19 - one-and-half column

Fig.20 – one column

Fig.21 - one-and-half column

Conflicts of interest statement

Manuscript title: Comparative Study of One Week-Ahead Forecasting of Daily Gas Consumption in Buildings Using Regression ARMA/SARMA and GA-optimized Regression Wavelet Neural Network Models

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Author names: Alexander Hošovský, Ján Piteľ, Milan Adámek, Jana Mižáková, Kamil Židek

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: