APPLICATION OF DEEP ESN NETWORKS AND WAVELET DECOMPOSITION FOR ENERGY CONSUMPTION FORECASTING IN BUILDINGS

DARYA LAPINA[1], ALEXANDER HOSOVSKY[1], JANA MIZAKOVA[1]

[1]Technical University of Kosice, Faculty of Manufacturing Technologies with the seat in Presov, Slovakia

DOI: 10.17973/MMSJ.2026_11_2026049

darya.lapina@tuke.sk

The aim of this study is to investigate the application of deep Echo State Networks (DeepESN) combined with wavelet decomposition for short-term forecasting of daily natural gas consumption in a building. The proposed approach includes extracting the temperature-related component using linear regression, wavelet decomposition of the residual signal, and independent forecasting of the resulting components using NAR, ESN, and DeepESN models. The experimental study was conducted using real data on natural gas consumption and outdoor air temperature from eleven heating seasons. The last season was used exclusively as an independent test set. Forecasting accuracy was evaluated using RMSE and MAE. The results showed that, among the investigated models, DeepESN provided the best forecasting performance for the complete time series, achieving an RMSE of 92.17 m^3/day and an MAE of 55.52 m^3/day. The results demonstrate the applicability of combining wavelet decomposition with DeepESN for short-term natural gas consumption forecasting.

KEYWORDS

Deep Echo State Networks, wavelet decomposition, natural gas consumption forecasting, time series forecasting, signal processing, short-term forecasting

1 INTRODUCTION

Energy is a fundamental basis of modern society, and buildings are among the significant energy consumers. The operation of buildings accounts for approximately 30% of global final energy consumption, with a significant share related to heating and hot water supply [IEA 2026]. Natural gas continues to play an important role in meeting the thermal needs of buildings, making the efficient management of its consumption a relevant task. Accurate forecasting of natural gas consumption is an important tool for improving energy efficiency and optimizing building operation. Reliable forecasts make it possible to plan energy consumption, optimize heating equipment operation, and reduce operating costs. However, forecasting gas consumption is a challenging task due to its dependence on temperature and other weather factors, seasonality, long-term trends, irregular fluctuations, and nonlinear dependencies [IEA 2026]. These characteristics limit classical linear methods and require models capable of accounting for complex temporal data structures. Statistical methods, machine learning methods, and neural networks are used for natural gas consumption forecasting. Statistical approaches include AR, MA, ARIMA, SARIMA, and SARIMAX models. They are relatively simple and interpretable; however, their assumptions of linearity and stationarity limit their ability to describe nonlinear dependencies and abrupt changes in consumption. The limitations of statistical approaches have also been observed in comparative studies of energy consumption forecasting [Wen 2020].

Among machine learning methods, Support Vector Regression (SVR) is widely used and is capable of modeling nonlinear dependencies using kernel functions. However, its effectiveness depends on the proper selection of input variables and time lags. In study [Wen 2020], neural models also showed lower MAPE values compared with ARIMA, SVM, and MLR, indicating the potential of neural network approaches for energy consumption forecasting. Neural networks are used for time series forecasting due to their ability to model complex nonlinear dependencies. The architectures used include multilayer perceptrons (MLP), recurrent networks, including LSTM, and nonlinear autoregressive networks (NAR). In study [Wen 2020], the best result among the considered models was demonstrated by DRNN-GRU with a MAPE of 3.504%, followed by DRNN-LSTM with 3.640%.

In recent years, deep learning architectures, including Transformer and Temporal Convolutional Networks (TCN), have been actively developing. Transformer uses an attention mechanism to account for dependencies between different elements of a sequence; however, its performance depends on the specific task and architecture. TCNs are used to model temporal dependencies; in study [Li 2022a], the TCN-StS model outperformed RNN and LSTM in forecasting accuracy across all considered forecasting horizons.

An alternative approach is Reservoir Computing (RC), which aims to simplify the training of recurrent neural networks. In the classical implementation of RC, the internal connections of the reservoir are fixed, while training is mainly limited to the output layer, reducing the computational complexity of training [Stepney 2024]. One of the most widely used representatives of RC is the Echo State Network (ESN), which uses a dynamic reservoir with fixed internal connections [Zhong 2017; Hu 2020]. The practical computational efficiency of ESN was demonstrated in [Stolzenburg 2025], where the average total training and testing time was 1.24 s compared with 125.43 s for LSTM.

An extension of ESN is the Deep Echo State Network (DeepESN), which uses a hierarchical organization of multiple reservoir layers [Hu 2020]. This structure enables sequential transformation of input information at multiple levels. In a comparative study on energy consumption forecasting, DeepESN showed lower RMSE and MAPE values compared with ESN and demonstrated better MAPE and sMAPE results on data from Canada and Japan [Hu 2020]. Similar results were presented in [Gallicchio 2019], where DeepESN was compared with ESN, SRN, LSTM, and GRU on four datasets and demonstrated the shortest training and testing time among the considered models while achieving higher accuracy compared with ESN.

These results provide a basis for considering DeepESN in the present study. The effectiveness of ESN and DeepESN depends on the selection of hyperparameters, whose optimal values may differ substantially depending on the specific forecasting task [Viehweg 2023]. Therefore, in the present study, the DeepESN hyperparameters are individually selected for the considered dataset using Bayesian optimization. In addition to model architecture selection, the representation of the original time series plays an important role. The complex structure of energy time series makes the application of signal decomposition methods relevant. One such tool is the Wavelet Transform (WT), which allows a signal to be analyzed in both the time and frequency domains and decomposed into components representing different characteristics of the original series [Joo 2015]. The practical effectiveness of WT in natural gas consumption forecasting was demonstrated in [Li 2022b], where

wavelet decomposition was applied to daily natural gas consumption and the resulting components were forecast using a BiLSTM-based model. The results showed that the choice of the wavelet type and decomposition level affected forecasting performance [Li 2022b]. Thus, the reviewed studies demonstrate the effectiveness of wavelet decomposition for representing the complex structure of time series and of ESN/DeepESN for modeling sequential data. However, their combined application to short-term natural gas consumption forecasting in buildings remains insufficiently studied. Such an approach is of additional interest when using daily data and a limited number of observations, where the choice of the time series representation and model architecture may have a substantial impact on forecasting performance. Based on the reviewed analysis, the aim of the present study is to investigate the effectiveness of combining wavelet decomposition and Deep Echo State Networks (DeepESN) for short-term forecasting of natural gas consumption in a building and to compare the obtained results with a conventional nonlinear autoregressive model (NAR) and a classical Echo State Network (ESN). To achieve this aim, the following tasks are addressed in the study:

1. construction of a linear regression model of natural gas consumption as a function of outdoor air temperature and obtaining the residual time series;
2. decomposition of the residual time series using the Wavelet Transform and determination of an appropriate decomposition level;
3. training and optimization of NAR, ESN, and DeepESN models for individual wavelet decomposition components;
4. evaluation and comparison of model accuracy on an independent test heating season and analysis of the obtained results.

2 DATA DESCRIPTION

The present study uses real data on the daily natural gas consumption of the administrative and medical building of the National Institute of Rheumatic Diseases in Piešťany, Slovakia. The data cover the period from August 15, 2014 to June 30, 2025 and contain 3974 daily observations. For each day, the total natural gas consumption over a 24-hour period was recorded. Additionally, the average daily outdoor air temperature was used, calculated based on measurements taken in the morning, at noon, and in the evening. Fig. 1 presents the original time series of gas consumption and temperature.

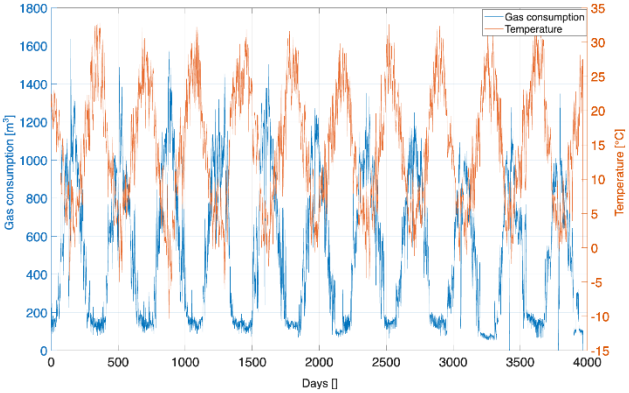


Figure 1. Time series of gas consumption and temperature

Gas consumption is characterized by pronounced seasonality and dependence on outdoor air temperature, which is mainly related to the use of gas for building heating. Therefore, the subsequent analysis considers heating seasons, defined as the period from October 1 to May 1. Ten complete heating seasons from 2014/2015 to 2023/2024 were used for model training. After excluding February 29 in leap years, 2130 daily observations were obtained. The following heating season, 2024/2025, was not used for model training or tuning. From this season, 177 daily observations were selected and used exclusively for independent testing of the obtained models.

Before modeling, the training set was checked for missing and anomalous values. Five missing values (0.25%) were identified and reconstructed using linear interpolation. Anomalous values, mainly associated with monitoring failures, maintenance work, or equipment replacement, were checked and, where necessary, corrected or excluded. The exclusion of February 29 ensured the same duration for all considered heating seasons.

3 METHODOLOGY

3.1 Regression-based decomposition

Simple linear regression describes the quantitative relationship between a dependent variable and an explanatory variable and can be represented as [Rotova 2020]:

$$y = a_0 + a_1 * x + \varepsilon \quad (1)$$

where y is the dependent variable, x is the explanatory variable, a_0 and a_1 are the model parameters, and ε is the residual component [Rotova 2020]. In the present study, daily natural gas consumption is used as the dependent variable, while the average daily outdoor air temperature is used as the explanatory variable. This choice reflects the influence of temperature on building heating demand and gas consumption. After constructing the model, the residual series is calculated as:

$$\varepsilon_t = y_t - \hat{y}_t \quad (2)$$

where $\hat{y}_t$ is the consumption value estimated by the regression model. The residuals represent unexplained consumption variation and may contain additional temporal patterns. In the present study, the residual series is used as an independent time series for further analysis. It is subjected to wavelet decomposition, after which the obtained components are forecast using NAR, ESN, and DeepESN models.

3.2 Wavelet decomposition

Wavelet decomposition is applied for preprocessing the residual series, allowing the signal to be represented in terms of components characterizing different time scales. The residual series is represented as the sum of an approximation component and detail components:

$$y(t) = A_J(t) + \sum_{j=1}^{J} D_j(t) \quad (3)$$

where $A_J(t)$ is the approximation component describing the long-term dynamics of the time series, and $D_j(t)$ are the detail components characterizing its fluctuations at different time scales j. Lower decomposition levels correspond to faster changes in the signal, whereas higher levels reflect slower variations.

The selection of the mother wavelet type and decomposition level is an important stage of wavelet analysis, as it affects the quality of the time series representation and the characteristics of the extracted components. In the present study, a wavelet from the Daubechies family is used. The choice of this family is motivated by its compact support, orthogonality, and vanishing moments [Liu 2018]. The Daubechies tenth-order wavelet (db10) is selected as the mother wavelet, providing a compromise between the smoothness of the representation and signal localization [Liu 2018].

The decomposition level determines the depth of the multilevel analysis, and its selection should take into account the length of the analyzed signal and the structure of the investigated time

series [Khan 2023]. The maximum allowable decomposition level is estimated using the empirical rule [Khan 2023]:

$$J_{max} = \text{fix}\left[\log_2\left(\frac{N}{N_w} - 1\right)\right] \quad (4)$$

where N is the length of the analyzed signal, N_w is the filter length of the selected mother wavelet function, and fix denotes taking the integer part of the obtained value [Khan 2023].

In the present study, the analyzed residual series contains (N=2130) observations, and the db10 wavelet with a filter length of (N_w=20) is used for decomposition. According to the above rule:

$$J_{max} = \text{fix}\left[\log_2\left(\frac{2130}{20} - 1\right)\right] = \text{fix}(6.72) = 6 \quad (5)$$

Thus, the sixth decomposition level is used, resulting in the residual series being represented by one approximation component A6 and six detail components D1, …, D6. The resulting seven components are treated as separate time series and are subsequently used to construct forecasting models.

3.3 Training

3.3.1 Training and validation data

The six detail components D1,..,D6 and the approximation component A6 obtained from the wavelet decomposition are used to train and compare three forecasting models: NAR, ESN, and DeepESN. A separate model is independently constructed for each of the seven wavelet components.

The first ten complete heating seasons, containing a total of (N=2130) daily observations, are used for model training. For each wavelet component, the resulting series of 2130 observations is divided into training and validation sets in an 80:20 ratio:

$$D = D_{train} \cup D_{val} \quad (6)$$

$$|D_{train}| = 0.8|D|\ ,\ |D_{val}| = 0.2|D|$$

The following heating season, 2024/2025, containing 177 daily observations, is completely excluded from training and optimization and is used exclusively as an independent test set for the final evaluation of forecasting performance.

3.3.2 Forecasting strategy

The present study uses a direct multi-step forecasting strategy, in which the value at a given horizon h is modeled directly without sequentially using previous forecasts. Thus, errors from previous forecasting steps are not propagated to subsequent forecasts [Green 2025]. The model input consists of available time series values at preceding time points, while the output is the value h time steps ahead:

$$X_t \rightarrow X_{\text{t+h}} \quad (7)$$

where h denotes the forecasting horizon. In the present study, the forecasting horizon is set to h =7 days.

3.3.3 Optimization criterion

Bayesian optimization is used for tuning the model hyperparameters in the present study. It selects subsequent configurations based on the results of previous evaluations, reducing the number of computational experiments required compared with exhaustive search [Koehrsen 2018].

The Root Mean Square Error (RMSE) is used as the main objective function for optimization:

$$RMSE = \sqrt{\frac{1}{N_{val}}\sum_{t=1}^{N_{val}}(y_t - \hat{y}_t)^2} \quad (8)$$

where N_{val} is the number of observations in the validation set, y_t is the actual value, and $\hat{y}_t$ is the forecasted value [Svetunkov 2023].

Thus, the hyperparameter optimization problem is formulated as:

$$\theta^* = \arg\min_{\theta\in\Theta} RMSE\ (\theta) \quad (9)$$

where θ is the vector of model hyperparameters and Θ is the space of their admissible values [Koehrsen 2018].

Mean Absolute Error (MAE) is used as an additional performance metric [Svetunkov 2023]:

$$MAE = \frac{1}{N_{val}}\sum_{t=1}^{N_{val}}|y_t - \hat{y}_t| \quad (10)$$

Thus, RMSE is used directly for hyperparameter optimization, while RMSE and MAE are used to evaluate the accuracy of the obtained forecasts. Lower values of both metrics correspond to a lower forecasting error.

3.3.4 Bayesian optimization

Bayesian optimization is applied independently to the NAR, ESN, and DeepESN models for each wavelet component, with validation RMSE as the objective function. The specific hyperparameters and their search ranges are defined in the corresponding model subsections.

3.3.5 NAR model

The nonlinear autoregressive model (Nonlinear Autoregressive, NAR) is used as the baseline model. It forecasts the future value of a time series based on its previous values [Katy 2025].

For a univariate time series, the general principle of the NAR model can be represented as follows:

$$y(t) = f\big(y(t-1), y(t-2), \dots, y(t-n)\big) + \varepsilon(t) \quad (11)$$

where $y(t)$ is the time series value at time t , $y(t-1), y(t-2), \dots, y(t-n)$ are the previous values of the series, n is the number of delays used, f is a nonlinear function approximated by a neural network, and $\varepsilon(t)$ is a random error [Katy 2025].

The NAR model structure includes input, hidden, and output layers, where the lagged values of the time series are transformed by the hidden layer and used to generate the forecasted value [Zhang 1998]. The NAR model architecture and the set of delays used are determined during the optimization process. The number of hidden layers n_l, the number of neurons n_h, and the maximum delay window L are considered as hyperparameters [Ghahari 2021]. The search space for the NAR architectural parameters is defined as:

$$n_l \in [1,3], n_h \in [5,30], L \in [1,50] \quad (12)$$

The maximum delay window is specified separately for each component as the upper bound of the set of potential delays. The specific value L is used as a discrete parameter during the subsequent delay selection.

The optimization of the architectural parameters is performed using the Bayesian optimization described above, with RMSE on the validation set used as the objective function. At this stage, the optimal NAR architecture configuration and the corresponding maximum delay window are determined for each wavelet component.

After determining the maximum delay window, a separate stage is performed to select the specific delays using Bayesian optimization with binary variables. Within the specified window L, each possible delay i is associated with a binary variable:

$$z = (z_1, z_2, \dots, z_L), z_i \in \{0,1\} \quad (13)$$

where $z_i = 1$ indicates that the corresponding delay is included in the model input vector, while $z_i = 0$ indicates its exclusion. This binary formulation allows the selection of a subset of elements to be represented as an optimization problem in a discrete combinatorial space [Baptista 2018]. Thus, the

maximum delay window itself is not optimized; rather, a specific subset of delays within the specified range is optimized.
The objective function of the binary optimization is also defined in terms of the validation RMSE:

$$z^* = \arg\min_{z\in\{0,1\}^L} f(z) = RMSE(z) \quad (14)$$

During the optimization, different combinations of binary variables are evaluated, and the results of previous evaluations are used to select subsequent configurations, corresponding to the general principle of Bayesian optimization for expensive black-box functions [Baptista 2018]. As a result, a set of delays providing the minimum validation error among the considered configurations is determined.NAR tuning is performed in two stages: standard Bayesian optimization determines the architectural parameters and the maximum delay window, after which binary Bayesian optimization selects the specific set of delays. The procedure is performed independently for each of the seven wavelet components. Based on the identified parameters, the final model is constructed and trained and, without further tuning, is used to forecast the corresponding component on the independent test set.

3.3.6 ESN model

The second considered model is the Echo State Network (ESN), which belongs to the Reservoir Computing class. In ESN, the internal structure of the reservoir remains fixed after initialization, while training is performed for the output layer [Jaeger 2001]. The reservoir state is determined by the current input and the previous state:

$$x(t) = \tanh(W_{in}u(t) + W_x(t-1)) \quad (15)$$

where $u(t)$ is the input signal, $x(t)$ is the reservoir state vector, W_{in} the input weight matrix, and W is the matrix of internal reservoir connections [Jaeger 2001].
ESN training is performed only for the output layer. The output weights are determined using regularized linear regression (ridge regression):

$$W_{out} = YX^T(XX^T + \lambda I)^{-1} \quad (16)$$

where X is the reservoir state matrix, Y is the target values, λ is the regularization parameter, and I is the identity matrix [Thiede 2019]. Thus, when tuning ESN, it is necessary to determine the hyperparameters characterizing the size and dynamic properties of the reservoir, as well as the regularization of the output layer. In the present study, four main parameters are optimized: reservoir size N, spectral radius of the internal connection matrix $p(W)$, input scaling coefficient a, and output layer regularization parameter λ [Thiede 2019]. The following hyperparameter ranges are defined as the search space:

- reservoir size: $N \in [50, 300]$
- spectral radius of the reservoir matrix: $p(W) \in$ [0.1, 1.2]
- input scaling: $a \in [0.1, 1.0]$
- output layer regularization parameter: $\lambda \in [10^{-6}, 10^{-2}]$

Optimization is performed using Bayesian optimization based on the RMSE of the validation set independently for each of the seven components. The subsequent procedure of training, saving the models, and applying them to the independent test set is performed analogously to that described for the NAR model.

3.3.7 DeepESN model

The Deep Echo State Network (DeepESN) is an extension of ESN with a multilayer organization of reservoirs. The state of the first layer is determined by the input signal, while each subsequent layer receives the state of the previous layer [Gallicchio 2017] [Gallicchio 2018]. Let the network consist of L reservoir layers. The state of the first layer is determined by the current input signal and its previous state [Gallicchio 2017]:

$$x^{(1)}(t) = (1-\beta)x^{(1)}(t-1) + \beta \tanh\left(W_{in}^{(1)}u(t) + W^{(1)}x^{(1)}(t-1)\right) \quad (17)$$

For the subsequent layers $l = 2, \dots, L$, the input is the state of the previous layer:

$$x^{(l)}(t) = (1-\beta)x^{(l)}(t-1) + \beta \tanh\left(W_{in}^{(l)}x^{(l-1)}(t) + W^{(l)}x^{(l)}(t-1)\right) \quad (18)$$

Thus, each layer maintains its own state while simultaneously using information received from the previous level [Gallicchio 2017].
The final DeepESN state is formed by concatenating the states of all reservoir layers:

$$x(t) = [x^{(1)}(t);\ x^{(2)}(t); \dots; x^{(L)}(t)] \quad (19)$$

after which the output layer generates the forecasted value:

$$\hat{y}(t) = W_{out}x(t) \quad (20)$$

DeepESN uses the same main hyperparameters as ESN, with the number of reservoir layers L added as a structural parameter. In the present study, five hyperparameters are optimized: reservoir size N,spectral radius of the internal connection matrix $p(W)$, input scaling coefficient a, output layer regularization parameter λ ,and number of reservoir layers L. Increasing the depth allows a sequence of dynamic representations of the time series to be formed at different levels of processing [Thiede 2019; Gallicchio 2017 & 2018]. The same search ranges as for the classical ESN are used for the first four hyperparameters. Additionally, the number of reservoir layers is set within the range $L \in [2, 4]$. The number of layers is limited to L∈[2,4] to control architectural complexity. All hyperparameters are optimized using Bayesian optimization based on the RMSE of the validation set independently for each of the seven components. The subsequent procedure of training, saving the models, and applying them to the independent test set is performed analogously to that described for the previous models.

3.4 Testing and signal reconstruction

As a result of the described training and optimization procedure, trained NAR, ESN, and DeepESN models were obtained for each of the seven wavelet components. The final evaluation was performed using an independent test set corresponding to the eleventh heating season and containing 177 daily observations. After the residual component was extracted using the regression model, the test signal was treated as a single time series for the analysis of its component structure. The components obtained through wavelet decomposition were used to generate forecasts with the corresponding pre-trained models. The result of processing the test signal is presented in Fig. 2.

Due to the selected forecasting horizon h=7, the first seven values of the test series do not have corresponding forecasts; therefore, the comparison between forecasted and actual values is performed on 170 observations. For each of the three models, the forecasts of the seven wavelet components are combined by summation, resulting in a forecast of the residual signal:

$$\hat{R}(t) = \sum_{i=1}^{6} \widehat{D_i}(t) + \widehat{A_6}(t) \quad (21)$$

where $\hat{R}(t)$ represents the reconstructed residual signal for the eleventh season.

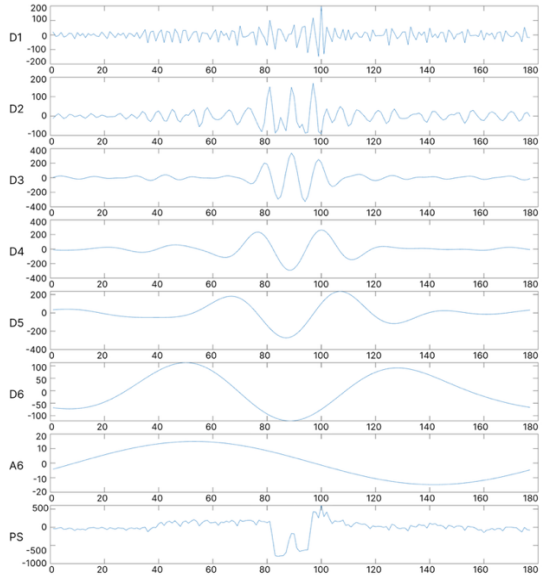


Figure 2. Wavelet decomposition of the residual signal for the eleventh heating season

The resulting residual forecast was combined with the temperature-related component calculated using the actual outdoor air temperature values to reconstruct the forecast of the original natural gas consumption:

$$\hat{y}(t) = \widehat{y_{reg}}(t) + \hat{R}(t) \tag{22}$$

where $\widehat{y_{reg}}(t)$ denotes the temperature-related component calculated using the actual outdoor air temperature values; $\hat{R}(t)$ represents the reconstructed forecast of the residual component; and $\hat{y}(t)$ denotes the final forecast of daily natural gas consumption. Thus, the final forecast was generated under the assumption that actual outdoor temperature values for the corresponding dates were available, rather than temperature forecasts. The resulting forecasts were compared with the actual natural gas consumption values for the corresponding 170 observations of the eleventh heating season. Forecasting accuracy was evaluated using the RMSE and MAE metrics. The overall scheme of the proposed forecasting procedure is presented in Fig. 3.

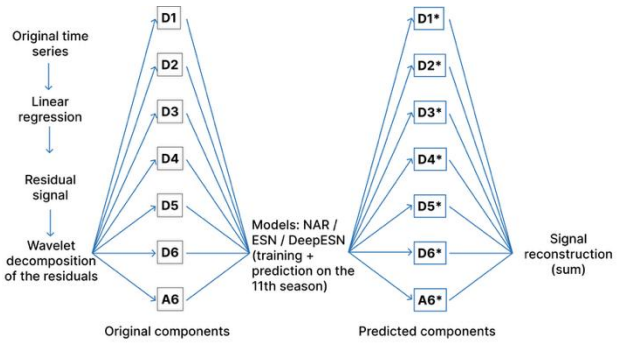


Figure 3. General scheme of the proposed forecasting procedure

All models are evaluated on the same independent test set under identical forecasting and optimization conditions, ensuring their comparability.

4 RESULTS AND DISCUSSION

This section presents the results of the conducted study. First, the results of the linear regression and wavelet decomposition of the residual time series are considered. Next, the forecasting accuracy of individual wavelet components is evaluated using NAR, ESN, and DeepESN models. Finally, the quality of the reconstructed forecasts of the original time series is compared on the independent test set.

4.1 Linear regression results

To quantitatively assess the relationship between daily natural gas consumption and outdoor air temperature, a linear regression model was constructed. The resulting equation is:

$$Q = -45.85 * T + 1155.23 \tag{23}$$

where Q is the daily natural gas consumption and T is the average daily outdoor air temperature. The negative value of the temperature coefficient indicates an inverse relationship between the considered variables: as the temperature decreases, natural gas consumption increases. The coefficient of a determination was $R^2 = 0.798$, which means that approximately 79.8% of the variation in daily natural gas consumption is explained by changes in outdoor temperature within the constructed linear model. This result confirms the substantial influence of temperature on gas consumption in the investigated building and justifies its use for extracting the temperature-related component of the time series. The obtained SSE and RMSE values were $(3.52 * 10^7)$ and 128.5532 m³/day, respectively. These indicators characterize the approximation error of the regression model, while the remaining part of the consumption variation is not explained by the linear dependence on temperature and is retained in the residual component of the time series. The main performance indicators of the constructed regression model are presented in Table 1.

Metric	Value
R^2	0,798
SSE	$3{,}52\times10^7$
RMSE	128.5532

Table 1. Values of Selected Accuracy Metrics of the Linear Regression Model

Fig. 4 shows the linear regression between daily gas consumption and average outdoor air temperature, and Fig. 5 presents the resulting residuals, which are subsequently used for wavelet decomposition and nonlinear forecasting.

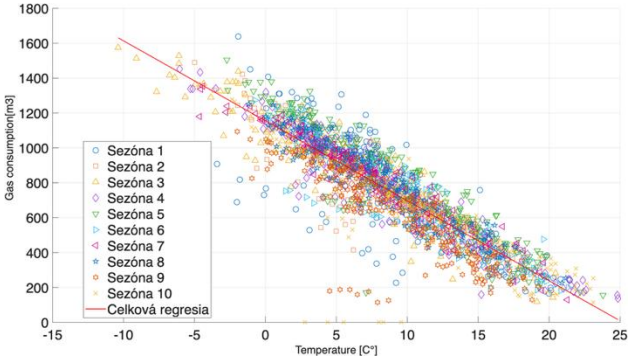


Figure 4. Regression analysis of gas consumption as a function of temperature

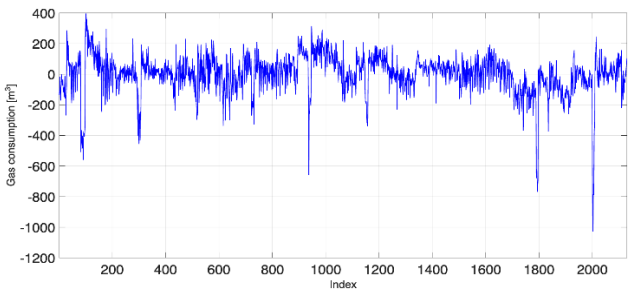


Figure 5. Residuals of the linear regression of the time series

4.2 Wavelet decomposition results

After extracting the residual component, wavelet decomposition was performed up to the sixth level. Fig. 6 presents the decomposition of the residual time series using the Daubechies 10 mother wavelet, including the original residual signal, the detail components D1-D6, and the approximation component A6. Visual analysis shows that the lower decomposition levels contain faster oscillations, whereas the components become smoother as the decomposition level increases. The approximation component A6 preserves the main slowly varying structure of the residual signal. All obtained components are subsequently used independently to construct NAR, ESN, and DeepESN models.

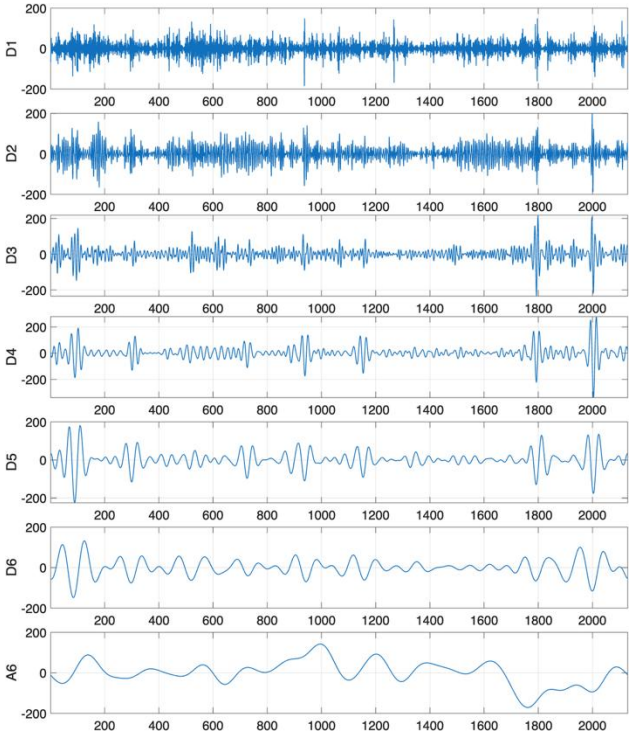


Figure 6. Wavelet decomposition of the residual time series

4.3 Model optimization and training results

At this stage of the study, the model hyperparameters are optimized, followed by model training using the obtained optimal configurations. For each wavelet component, optimization is performed independently using the Bayesian optimization procedure described above. For NAR, binary Bayesian optimization is additionally applied to select informative lags from the specified maximum delay window.The optimal hyperparameter values obtained from the optimization procedures for the NAR, ESN, and DeepESN models are presented in Table 2. For the NAR model, the parameters are denoted as neurons, layers, and maximum lag (N,L,ML). For the ESN and DeepESN models, the parameters are reservoir size, spectral radius, input scaling, and regularization parameter (RS, SR, IS, and RP), while DeepESN additionally includes the number of reservoir layers (RL).

Based on the optimal hyperparameters presented in Tab. 2, NAR, ESN, and DeepESN models were configured and trained for each of the seven wavelet components. Tab. 3 presents the RMSE and MAE values obtained for the trained models using the optimized hyperparameters. The presented indicators characterize the accuracy of the models for individual wavelet components and allow the performance of NAR, ESN, and DeepESN to be compared when modeling different time scales of the residual signal.

	NAR (N, L, ML)	ESN (RS, SR, IS, RP)	DeepESN (RS, SR, IS, RP, RL)
D1	5, 3, 50	52, 0.1, 0.918, 0.00637	78, 0.137, 0.118, 0.0012161, 4
D2	5, 1, 50	71, 0.98786, 0.13844, 1.3508e-06	174, 0.1688, 0.11697, 0.0098205, 4
D3	30, 2, 40	210, 0.88466, 0.13216, 0.0025803	241, 0.10051, 0.13617, 0.0013375, 4
D4	16, 3, 40	200, 0.4317, 0.10592, 0.0034132	132, 0.13278, 0.10273, 4.6644e-05, 3
D5	5, 3, 30	299, 0.22187, 0.10868, 0.00060114	226, 0.61535, 0.10212, 0.0011836, 3
D6	27, 3, 30	78, 0.1429, 0.10134, 0.00016213	176, 0.13983, 0.12668, 0.00056971, 4
A6	18, 3, 30	240, 0.5357, 0.74936, 0.0049585	56, 0.11326, 0.13552, 0.0080468, 4

Table 2. Optimal hyperparameter values of the NAR, ESN, and DeepESN models

	NAR RMSE/MAE	ESN RMSE/MAE	DeepESN RMSE/MAE
D1	46.16/32.840	35.970/25.833	36.150/25.764
D2	38.701/27.958	40.805/29.335	**37.708/27.086**
D3	49.746/33.745	**31.644/20.354**	32.364/21.320
D4	39.347/24.713	46.462/26.448	**28.165/15.401**
D5	28.104/5.726	15.108/9.024	**6.095/4.105**
D6	11.253/3.610	7.496/5.082	**2.622/1.900**
A6	21.406/13.302	7.997/5.482	**6.839/3.495**

Table 3. RMSE and MAE values of the trained models for individual wavelet components

The comparison of RMSE and MAE shows differences in model performance when forecasting individual wavelet components. For D2, D4, D5, D6, and A6, the DeepESN model provides the lowest values of both metrics, whereas for D3, the lowest errors are obtained with ESN. For D1, the results are close: ESN has a slightly lower RMSE, while DeepESN has a slightly lower MAE. Thus, DeepESN shows the lowest errors for most components; however, the performance advantage of the models depends on the specific component.A tendency toward lower errors is also observed for components at higher decomposition levels, particularly D5, D6, and A6. This may be related to their slower and smoother variations, whereas the lower-level components contain faster oscillations and local changes that are more difficult to forecast. However, the relationship is not strictly monotonic and is determined by the characteristics of the

dynamics of each component.The obtained results characterize the forecasting performance of individual components; the final evaluation is performed after reconstructing the complete signal on the independent test set.

4.4 Testing results for individual wavelet components

To further evaluate the generalization ability of the trained models, they are tested on an independent dataset represented by the eleventh heating season. At this stage, the forecasting performance of each of the seven wavelet components D1-D6 and A6 is evaluated separately. RMSE and MAE values are calculated for each component and each model.

	NAR RMSE/MAE	ESN RMSE/MAE	DeepESN RMSE/MAE
D1	60.769/37.801	**39.600/27.762**	40.133/28.794
D2	51.290/34.846	**36.574/24.084**	45.473/25.597
D3	78.652/41.151	71.347/34.969	56.528/29.899
D4	45.279/30.174	54.065/35.473	**28.112/20.055**
D5	79.404/33.441	41.053/28.400	**17.061/12.868**
D6	23.593/10.690	13.320/9.801	**3.807/2.735**
A6	9.418/3.669	5.964/1.317	**1.360/0.603**

Table 4. RMSE and MAE values of the models for individual wavelet components on the independent test set

The independent testing results show noticeable differences in the forecasting accuracy of individual wavelet components for the NAR, ESN, and DeepESN models. DeepESN provides the lowest RMSE and MAE values for five of the seven considered components: D3, D4, D5, D6 and A6. For D1 and D2, lower values of both metrics are obtained with ESN. NAR is characterized by the highest error values in most cases, particularly for components D1, D2, D4, and D5. For component D1, RMSE decreases from 60.77 for NAR to 39.60 for ESN and 40.13 for DeepESN. A similar trend is observed for D2. In both cases, the classical ESN provides the lowest values of both metrics, whereas the advantage of DeepESN becomes apparent starting from higher decomposition levels.Particular attention should be paid to component D3, for which some of the highest forecasting errors are obtained. RMSE is 78.65 for NAR, 71.35 for ESN, and 56.53 for DeepESN, while MAE is 41.15, 34.97, and 29.90, respectively. Thus, despite the increased difficulty of forecasting this component, DeepESN provides a noticeable reduction in error compared with NAR and ESN. At the same time, D3 is characterized by the highest MAE among all components for each of the three models, indicating substantial deviations of the forecasts from the actual signal.For D4 and D5, the advantage of DeepESN becomes even more pronounced. For D5, the difference is particularly substantial: the DeepESN RMSE is 17.06, whereas values of 79.40 and 41.05 are obtained for NAR and ESN, respectively. A similar trend is observed for MAE, indicating a more accurate reproduction of the dynamics of these components by the DeepESN model. The lowest error values are observed for D6 and A6. For D6, RMSE is 23.59, 13.32, and 3.81 for NAR, ESN, and DeepESN, respectively, while for A6, the corresponding values are 9.42, 5.96, and 1.36. Thus, the approximation component A6 is the most predictable component for all three models. The low error values for the higher decomposition levels are generally consistent with their smoother and more slowly varying structure.

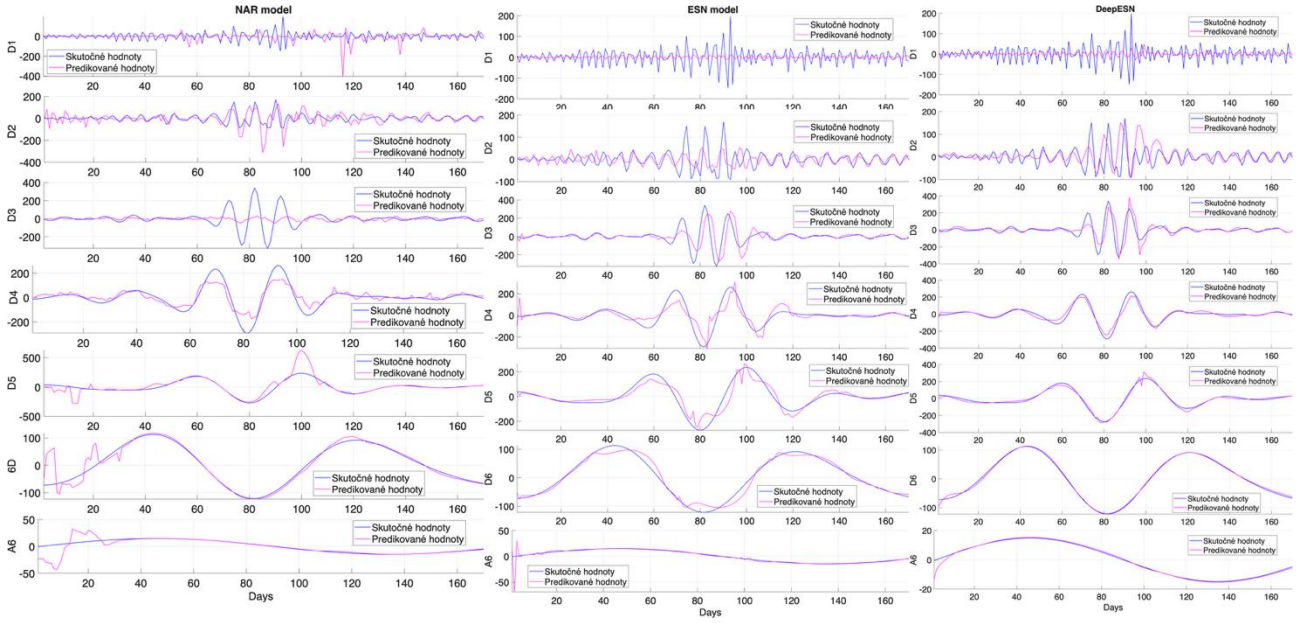


Figure 7. Comparison of the original and reconstructed time series for each wavelet component using NAR, ESN, and DeepESN

Visual analysis of the presented time series in Fig. 7 generally confirms the obtained RMSE and MAE values. For NAR, noticeable deviations from the actual values are observed in several cases, particularly around local peaks and rapid changes in the signal. In some sections, the model also produces oscillations that are absent from the actual time series. ESN forecasts generally correspond more closely to the actual dynamics; however, deviations in amplitude and local temporal shifts are also observed for individual components. In most cases, DeepESN most closely reproduces the structure of the actual signals, although small deviations and delays are also observed for this model in some sections. Overall, the obtained results show that forecasting difficulty depends on the time scale of the corresponding wavelet component. The highest errors are observed mainly for components D3-D5, whereas D6 and A6 are forecast considerably more accurately. At the same time, model selection also has a substantial effect on accuracy: DeepESN provides the lowest values of both metrics for five of the seven

components, ESN provides the lowest values for the two lower-level components D1 and D2, while NAR shows higher errors in most cases.

4.5 Results of Reconstruction of the Original Time Series

To reconstruct the complete time series, the forecasts of the seven wavelet components and the temperature-dependent component of the linear regression are used. It should be noted that the final comparison is not performed for all 177 observations of the eleventh heating season, but for 170 observations. The first seven values are excluded from the comparison due to the use of the forecasting horizon (h = 7), since the corresponding forecast values are not available for them. Similarly, the first seven observations are excluded from the original signal and the temperature-dependent component. Fig. 8 presents the original time series of the eleventh heating season and its shortened version used for the subsequent comparison with the reconstructed forecasts.

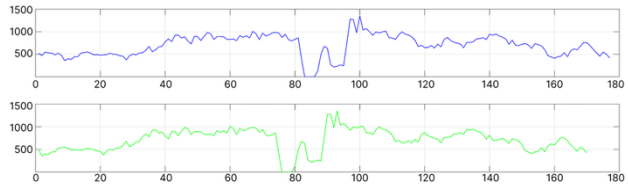


Figure 8. Original and shortened time series of the eleventh heating season

In accordance with the approach described in the methodology section, the forecasts of all wavelet components are summed to reconstruct the residual component, to which the temperature-dependent component obtained from the linear regression is then added. The resulting complete reconstructed signal is used for further evaluation against the original time series. At this stage, the main part of the analysis is performed, summarizing the results of the conducted study. The complete reconstruction makes it possible to evaluate the result of applying the entire proposed "decomposition + forecasting" scheme as a whole. The main quantitative results are presented in Table 5 as RMSE and MAE values for each of the considered models. A visual analysis of the obtained time series is then performed.

	NAR	ESN	DeepESN
RMSE/ MAE	160.729/ 102.210	138.704/ 78.064	92.171/ 55.518

Table 5. Comparison of RMSE and MAE Metrics for the Complete Reconstructed Signal (Trend + Residuals) Using NAR, ESN, and DeepESN Models

The obtained results show noticeable differences in the quality of full signal reconstruction between the considered models. The results are consistent with the observations obtained during independent testing of individual wavelet components, where DeepESN demonstrated the best RMSE and MAE values for five out of seven components. This trend is preserved when reconstructing the complete time series. The highest error values were obtained for the NAR model: RMSE = 160.73 and MAE = 102.21. For ESN, the values decrease to RMSE = 138.70 and MAE = 78.06, indicating more accurate reconstruction of the complete signal compared with NAR. The lowest error values were obtained for DeepESN: RMSE = 92.17 and MAE = 55.52. The visual analysis also shows differences between the models. NAR is characterized by the most noticeable deviations from the original signal, especially in the areas of local peaks and abrupt changes. For ESN, a closer correspondence to the original signal is observed, although some deviations remain, mainly during abrupt changes. DeepESN reproduces the overall structure of the original time series and the main peaks most closely, although some smoothing of extreme values is observed in certain sections. Thus, the visual comparison is generally consistent with the obtained RMSE and MAE values and confirms the differences between the models identified through quantitative evaluation. For a more detailed analysis of the nature of the errors, their distribution is further examined using boxplot graphs and histograms. Unlike RMSE and MAE, these representations additionally make it possible to assess the spread of errors, the presence of outliers, and the stability of the obtained results across different sections of the time series.

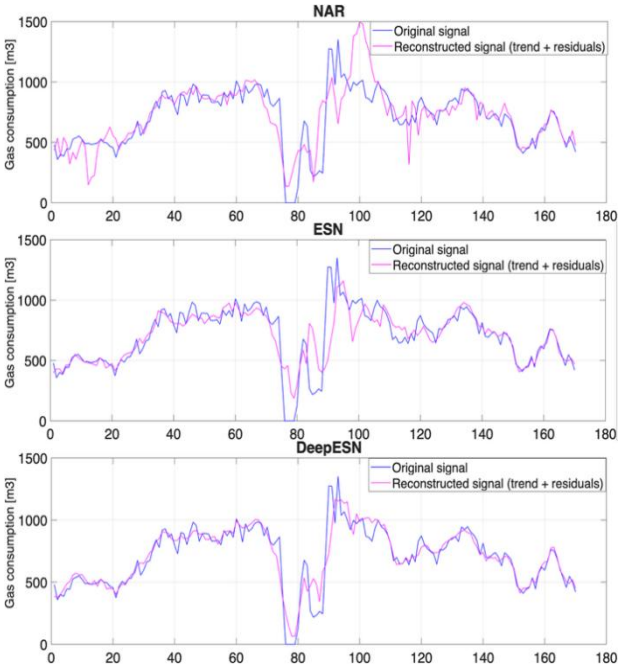


Figure 9. Comparison of the Reconstructed and Actual Signals of the 11th Heating Season for NAR, ESN, and DeepESN

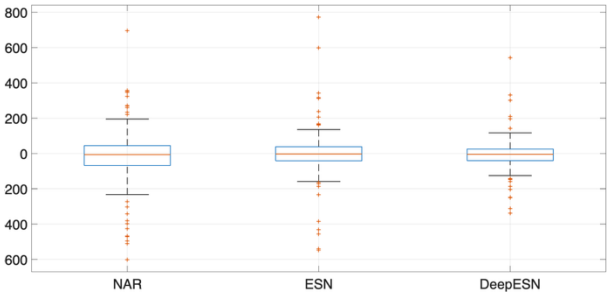


Figure 10. Boxplot Analysis of the Error Distribution for the NAR, ESN, and DeepESN Models

The boxplot analysis provides an additional assessment of the distribution of reconstruction errors for the considered models, including their spread and the presence of individual significant deviations. For NAR, the widest interquartile range is observed, indicating high variability of errors across different sections of the time series. In addition, a larger number of outliers is observed for this model, indicating the presence of individual sections with substantially higher reconstruction errors. For ESN, the error distribution is more compact compared with NAR. The reduction in the interquartile range and the smaller number of outliers indicate a more uniform distribution of errors, although some significant deviations remain. The most compact error distribution is observed for DeepESN. The interquartile range is the smallest among the considered models, and the number of outliers is also lower. This indicates a smaller spread of errors and more uniform reconstruction quality throughout the test period.

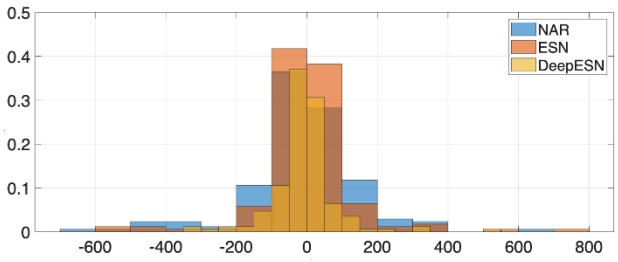


Figure 11. Histograms of the Error Distribution for the Reconstructed Complete Signal Using NAR, ESN, and DeepESN

The histograms further characterize the distribution of reconstruction errors for the considered models. NAR shows the widest error distribution and the lowest concentration around zero, indicating greater variability throughout the test period. For ESN, the distribution is more compact, with values concentrated in the central region, although some deviations remain. DeepESN exhibits the strongest concentration of errors around zero and a narrower main distribution, although some values remain more distant. These findings are consistent with the quantitative evaluation and boxplot analysis: DeepESN shows the most compact error distribution, NAR the highest variability, and ESN an intermediate performance. Overall, the analysis of individual wavelet components, the reconstructed signal, and error distributions provides a comprehensive assessment of the models, highlighting differences in both reconstruction accuracy and error characteristics.

5 CONCLUSIONS

This study investigated an approach to short-term forecasting of daily natural gas consumption based on the extraction of the temperature-dependent component, wavelet decomposition of the residual signal, and forecasting of the resulting components using NAR, ESN, and DeepESN models. The conducted study showed that wavelet decomposition is a suitable stage of preprocessing the residual signal before forecasting. Separating the signal into components with different temporal characteristics made it possible to identify differences in forecasting complexity and evaluate the performance of each model at individual decomposition levels. The results for individual components also showed that forecasting accuracy depends not only on the model used but also on the characteristics of the specific component. The main finding of the study is that DeepESN demonstrated the best forecasting performance for the task considered. On the independent test set, after reconstructing the complete time series, the DeepESN model achieved an RMSE of 92.17 and an MAE of 55.52, whereas the corresponding values for ESN were 138.70 and 78.06, and those for NAR were 160.73 and 102.21, respectively. DeepESN also achieved the lowest values for both error metrics on five out of seven individual wavelet components. Thus, among the models investigated, DeepESN provided the most accurate reconstruction of the complete natural gas consumption signal. Another practical observation was the substantial difference in model training times. NAR required considerably more time to train, whereas ESN and DeepESN were trained significantly faster. Although computational time was not measured as part of a separate quantitative experiment, this observation indicates an additional practical advantage of reservoir-based models for the task considered. Overall, the objective of the study was achieved: the proposed scheme of temperature-dependent component extraction, wavelet decomposition, component forecasting, and complete signal reconstruction demonstrated its applicability to short-term natural gas consumption forecasting. The results show that combining wavelet decomposition with DeepESN provides more accurate forecasts than the NAR and ESN models considered, establishing a basis for further investigation of the forecasting framework. Further research should focus on a more detailed investigation of wavelet decomposition level selection, including different criteria and approaches for determining the optimal level based on subsequent forecasting performance, as well as on alternative decomposition methods. Forecasting accuracy could be further improved by incorporating additional factors affecting natural gas consumption, such as meteorological and calendar variables and building characteristics. The use of forecasted meteorological data instead of actual temperature values and the transition to adaptive and online model learning are also promising directions for future research. In addition, future studies should investigate the forecasting scheme using a strictly sequential evaluation procedure in which only information available at each forecast origin is used to construct the wavelet components and generate predictions. This would provide a more realistic assessment of the framework's forecasting performance and help determine its potential for practical application.

ACKNOWLEDGMENTS

This research was funded by the European Union under the Program Slovensko, project No. 401101C504; the Scientific Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic and the Slovak Academy of Sciences under the project VEGA 1/0061/23; and the Cultural and Educational Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic under the project KEGA 063TUKE-4/2025.

REFERENCES

[Baptista 2018] Baptista, R. and Poloczek, M. Bayesian Optimization of Combinatorial Structures. Proceedings of Machine Learning Research, 2018, Vol. 80, pp. 462-471.

[Gallicchio 2017] Gallicchio, C., et al. Deep reservoir computing: A critical experimental analysis. Neurocomputing, 2017, Vol. 268, pp. 87-99. ISSN 0925-2312.

[Gallicchio 2018] Gallicchio, C., et al. Design of deep echo state networks. Neural Networks, 2018, Vol. 108, pp. 33-47. ISSN 0893-6080.

[Gallicchio 2019] Gallicchio, C., et al. Comparison between DeepESNs and gated RNNs on multivariate time-series prediction. In: ESANN 2019 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019. ISBN 978-287-587-065-0. DOI: 10.48550/arXiv.1812.11527

[Ghahari 2021] Ghahari, S., et al. Cluster Forecasting of Corruption Using Nonlinear Autoregressive Models with Exogenous Variables (NARX)—An Artificial Neural Network Analysis. Sustainability, 2021, Vol. 13, No. 20, 11366. ISSN 2071-1050.

[Green 2025] Green, R., et al. Stratify: unifying multi-step forecasting strategies. Data Mining and Knowledge Discovery, 2025, Vol. 39, No. 5, 64. ISSN 1573-756X.

[Hu 2020] Hu, H., et al. Forecasting energy consumption and wind power generation using deep echo state network. Renewable Energy, 2020, Vol. 154, pp. 598-613. ISSN 0960-1481.

[IEA 2026] IEA. Buildings [online]. IEA, 2026 [cit. 2026-03-02]. Available from <https://www.iea.org/energy-system/buildings>

[Jaeger 2001] Jaeger, H. The "echo state" approach to analysing and training recurrent neural networks. GMD Report 148. Bonn: German National Research Center for Information Technology, 2001.

[Joo 2015] Joo, T. W. and Kim, S. B. Time series forecasting based on wavelet filtering. Expert Systems with Applications, 2015, Vol. 42, No. 8, pp. 3868-3874. ISSN 0957-4174.

[Katy 2025] Katy. Using Python to Implement Nonlinear Autoregressive Models (NAR) [online]. Dev Genius, 2025 [cit. 2026-01-27]. Available from <https://blog.devgenius.io/using-python-to-implement-nonlinear-autoregressive-models-nar-5ff7f2e867ef>

[Khan 2023] Khan, M.A., et al. Optimizing the Performance of Single-Phase Photovoltaic Inverter using Wavelet-Fuzzy Controller. e-Prime - Advances in Electrical Engineering, Electronics and Energy, 2022, Vol. 3, 100093. ISSN 2772-6711.

[Koehrsen 2018] Koehrsen, W. A Conceptual Explanation of Bayesian Hyperparameter Optimization for Machine Learning. Towards Data Science [online], 2018 [cit. 2026-02-18]. Available from <https://medium.com/towards-data-science/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f>

[Li 2022a] Li, H. and Qiu, T. Continuous Manufacturing Process Sequential Prediction using Temporal Convolutional Network. In: Proc. of 14th Int. Symp. on Process Systems Engineering – PSE, Kyoto, Japan, 19-23 June 2022. Elsevier B.V., 2022.

[Li 2022b] Li, Z., Liu, L., Qiao, W. Short-term natural gas consumption prediction based on wavelet transform and bidirectional long short-term memory optimized by Bayesian network. Energy Science & Engineering, 2022, Vol. 10, No. 9, pp. 3281–3300.

[Liu 2018] Liu, Y., et al. Wavelet-based 3-D inversion for frequency-domain airborne EM data. Geophysical Journal International, 2017, Vol. 213, No. 1, pp. 1-15. ISSN 0956-540X.

[Rotova 2020] Rotova, O.M. and Shibanova, A.D. Basic principles of the linear regression method [online]. Teoria i praktika sovremennoj nauki, 2020, No. 1 (55) [cit. 2026-01-18]. Available from https://cyberleninka.ru/article/n/osnovnye-printsipy-metoda-lineynoy-regressii/viewer (in Russian).

[Stepney 2024] Stepney, S. Physical reservoir computing: a tutorial. Natural Computing, 2024, Vol. 23, pp. 665-685. ISSN 1567-7818.

[Stolzenburg 2025] Stolzenburg, F., et al. Efficient time-series approximation with linear recurrent neural networks: architecture learning and predictive power. Neural Computing and Applications, 2025, Vol. 37, No. 32, pp. 27027-27055. ISSN 0941-0643.

[Svetunkov 2023] Svetunkov, I. Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM). Boca Raton: Chapman and Hall/CRC, 2023. ISBN 9781032590370.

[Thiede 2019] Thiede, L.A. and Parlitz, U. Gradient based hyperparameter optimization in Echo State Networks. Neural Networks, 2019, Vol. 115, pp. 23-29. ISSN 0893-6080.

[Viehweg 2023] Viehweg, J., et al. Parameterizing echo state networks for multi-step time series prediction. Neurocomputing, 2022, Vol. 522, pp. 214-228. ISSN 0925-2312.

[Wen 2020] Wen, L., Zhou, K., Yang, S. Load demand forecasting of residential buildings using a deep learning model. Electric Power Systems Research, 2020, Vol. 179, 106073. ISSN 0378-7796.

[Zhang 1998] Zhang, G., et al. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 1998, Vol. 14, No. 1, pp. 35-62. ISSN 0169-2070.

[Zhong 2017] Zhong, S., et al. Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. Neurocomputing, 2017, Vol. 238, pp. 191-204. ISSN 0925-2312.

CONTACTS:

Ing. Darya Lapina
Prof. Ing. Alexander Hosovsky, PhD.
PaedDr. Jana Mizakova, PhD.
Technical University of Kosice
Faculty of Manufacturing Technologies with the seat in Presov
Bayerova 1, 080 01 Presov, Slovakia
+421 055 602 6420
e-mail: darya.lapina@tuke.sk, alexander.hosovsky@tuke.sk, jana.mizakova@tuke.sk

LICENSE CREATIVE COMMONS:

The article is published under the terms and conditions of the Creative Commons Attribution 4.0 International License (CC BY 4.0).