

Article

Hybrid Load Forecasting Using Gaussian Process Regression and Novel Residual Prediction

Cosmin Darab ^{1*}, Antoniu Turcu ¹, Horia Gheorghe Beleiu ¹, Sorin Pavel ¹, Iulian Birou ², Dan Doru Micu ³, Stefan Ungureanu ¹ and Stefan Dragos Cirstea ¹

¹ Department of Electrical Engineering, Technical University of Cluj-Napoca, 28 Memorandumului Street, 400114 Cluj-Napoca, Romania; antoniuturcu@enm.utcluj.ro (A.T.); horia.beleiu@enm.utcluj.ro (H.G.B.); sorinpavel@enm.utcluj.ro (S.P.); stefan.ungureanu@enm.utcluj.ro (S.U.); stefan.cirstea@enm.utcluj.ro (S.D.C.)

² Department of Electric Machines and Drives, Faculty of Electrical Engineering, Technical University of Cluj-Napoca, 28 Memorandumului Street, 400114 Cluj-Napoca, Romania; iulian.birou@emd.utcluj.ro

³ Department of Electrotechnics and Measurements, Faculty of Electrical Engineering, Technical University of Cluj-Napoca, 28 Memorandumului Street, 400114 Cluj-Napoca, Romania; Dan.Micu@ethm.utcluj.ro

* Correspondence: cosmin.darab@enm.utcluj.ro

Received: 12 June 2020; Accepted: 29 June 2020; Published: 2 July 2020

Abstract: Short-term electricity load forecasting has attracted considerable attention as a result of the crucial role that it plays in power systems and electricity markets. This paper presents a novel hybrid forecasting method that combines an autoregressive model with Gaussian process regression. Mixed-user, hourly, historical data are used to train, validate, and evaluate the model. The empirical wavelet transform was used to preprocess the data. Among the perturbing factors, the most influential predictors that were recorded were the weather factors and day type. The developed methodology is upgraded using a novel closed-loop algorithm that uses the forecasting values and influential factors to predict the residuals. Most performance indicators that are computed indicate that forecasting the residuals actually improves the method's precision, decreasing the mean absolute percentage error from 5.04% to 4.28%. Measured data are used to validate the effectiveness of the presented approach, making it a suitable tool for use in load forecasting by utility companies.

Keywords: energy forecasting; Gaussian process regression; machine learning models; empirical wavelet transform; residuals prediction

1. Introduction

In recent years, considerable effort has been invested in the reduction of energy use, the promotion of renewable energy sources (RES), and the implementation of mitigation measures that will lead to a reduction in the environmental cost. The industrial sector is believed to account for 38% of total global energy consumption. Studies have shown that by simply implementing the energy efficiency solutions that are currently available, by 2040, industry could produce almost double the value per unit of energy, compared to its current levels [1]. Considering the emerging trends in energy consumption coupled with rising energy prices, it is crucial that the implementation of new forecast methodology is customized to users' specific profiles and capabilities of user profile.

Energy load forecasting has been one of the main focal points of energy-related research in both industry and academia in recent years. Nowadays, almost all decisions regarding the electricity market and utility industry use load forecasting as their primary criterion. For example, system operators use energy load forecasting to ensure power system reliability. It has also been proven to be a useful dynamical method for aggregators participating in the energy trading market or managing electricity consumption. Forecasting methods may essentially be classified as follows: very

short-term load forecasting (a few minutes ahead to a several hours), short-term (one day ahead to one week ahead), medium-term load forecasting (ranging from one month to one year), and long-term load forecasting (more than one year) [2]. Short-term predictions are used in demand and supply management and to balance electricity generation [3], transmission, distribution and use on different scales [4].

Another aspect of load forecasting in the context of the energy sector is the integration of RES into the power grid [5]. The use of a high-quality forecasting model may increase the impact of renewable sources [6]. In terms of energy management, load forecasting is an indispensable tool in demand response (DR) strategies [7]. However, the execution of reliable energy load forecasting is not easy, and several challenges arise in the process [8]. For this reason, researchers focus on overcoming these obstacles using new technologies, such as big data [9].

When energy load forecasting was still nascent, simple methods, such as linear regression [10], hybrid regressions [11], Kalman filters [12], Kalman estimators [13], or autoregressive integrating moving average (ARIMA) [14] and metabolic nonlinear grey models [15], were used. While these models are effective when dealing with linear data, they are unable to provide accurate forecasting in complex, non-linear energy load time series. Load forecasting methods have changed in response to advances in artificial intelligence neural network development [16]. Paras et al. used a neural networks approach for very short energy forecasting [17] and Nima et al. integrated a feature selection to the neural network algorithm [18]. As these models used optimal parameters [19], hybrid methods—basically, artificial neural network optimized methods—were proposed [20]. Many of the optimization methods used rely on feature selection [21] or hybrid feature selection to improve the parameters used for forecasting methodology [22]. Recent research has shown that performing a feature selection when starting to train the algorithm yields better results, as the model will be more likely to learn the characteristics of the time series when dealing with the optimum input-output set [23].

Hybrid algorithms that combine at least two prediction methods when developing the forecasting model have been developed [24]. Huaiguang et al. proposed a support vector regression method enforced with feature selection of data [25]. Sibonelo Motepe also proposed a hybrid methodology but using deep learning algorithms [26]. Due to the proven efficiency of these methods, the research presented herein is also based on a hybrid approach.

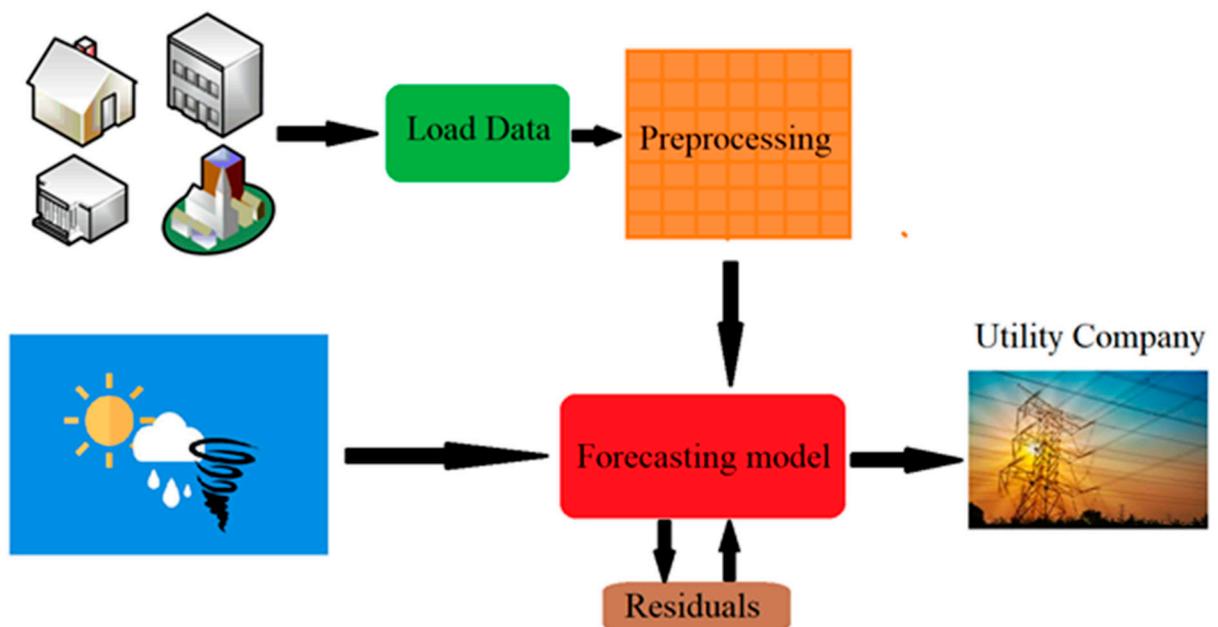


Figure 1. Schematic flow of the proposed method.

Unlike traditional forecasting methods that are used to predict the load on a single node (e.g., a substation), the focus of this paper is on short-term total energy load forecasting, which yields greater accuracy by using the measured information in all nodes of interest. Another benefit of the proposed methodology over the classical methods mentioned above is the use of the residuals as predictors. This approach will customize the algorithm to better respond to the characteristics of the data. By doing this with each iteration the algorithm will better adapt to the data sets.

The hybrid algorithm described in Figure 1, uses the autoregressive (AR) time series model in combination with Gaussian process regression (GPR), coupled with a data pre-processing algorithm, the empirical wavelet transform (EWT), for data optimization. This combination of parametric and non-parametric time series, AR and GPR, showed great results in predicting the energy load consumption characteristics [27]. For optimizing the forecasting model, the use of rational quadratic (RQ) kernel function is proposed. This covariance function was chosen after testing different other kernels, due to its compatibility with the non-linear profile of load data.

The remainder of this paper is organized as follows: Section 2 presents the energy load data used to develop the forecasting method. The subsequent section deals with data preprocessing and presents the EWT used. Section 4 presents the GPR hybrid forecasting method proposed in this paper. Subsequently, the model performance indicators used to assess the algorithm are described in Section 5. Section 6 presents the results and evaluates the forecasting method's performance. The final section of the paper concludes the study with some of the authors' observations and recommendations for future research regarding the presented topic.

2. Data Analysis

This paper tests how different factors impact energy consumption behavior. To develop the forecasting algorithms, load data were gathered from the Romanian energy transmission and system operator [28] and correlated with weather measurements for the city of Bucharest [29]. The weather data were correlated with energy load data to provide information at hourly intervals on the same time samples. Energy load data consist of energy consumption for every hour, starting with January 1, 2018 and ending on June 1, 2019. The weather data that were taken into account consist of five variables that were sampled in the same time intervals. The weather factors that were considered to have the greatest impact on load consumption were temperature, changes in atmospheric pressure, dew point temperature, the level of precipitation in mm, and air humidity, all measured at two meters above ground reference. Another important factor considered was whether or not the day was a working day. This data type is binary and is 1 if the day is a weekend day or a national holiday and is 0 otherwise. The final factor to be considered was the date (month, day, day of the week); this criterion was added as, upon data analysis, this principle was observed to create patterns in load demand. Prior to algorithm implementation, all data were processed, and an outlier test was performed to improve the datasets.

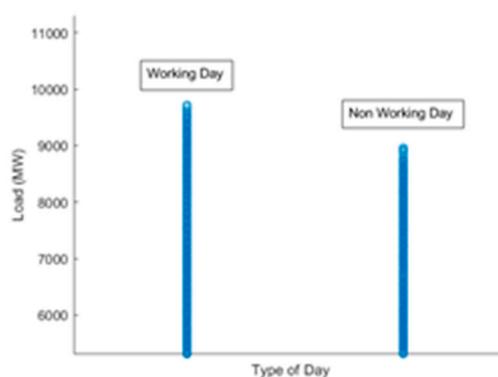


Figure 2. Energy consumption based on day type.

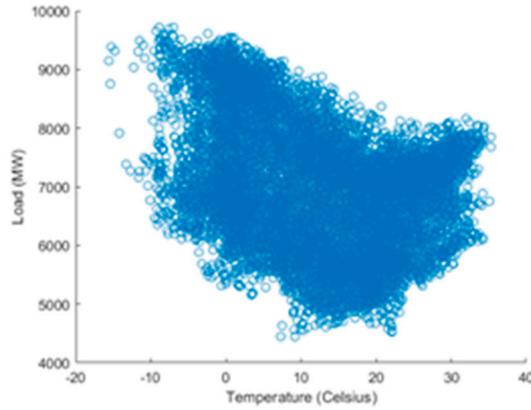


Figure 3. Temperature influence on energy load.

Figure 2 illustrates how energy load fluctuates in accordance with day type; as expected a significant increase in energy load for working days could be observed. Figure 3 represents temperature influence on load demand. A clear dependency of the load with respect to the temperature could be observed. Further interesting pattern-forming characteristics for the datasets are presented in Figures 4 and 5, which illustrate the influences of humidity and dew point temperature on power consumption, respectively. These characteristics are studied in order to determine if a feature selection would benefit the algorithm and which of the predictors influence less the energy load data.

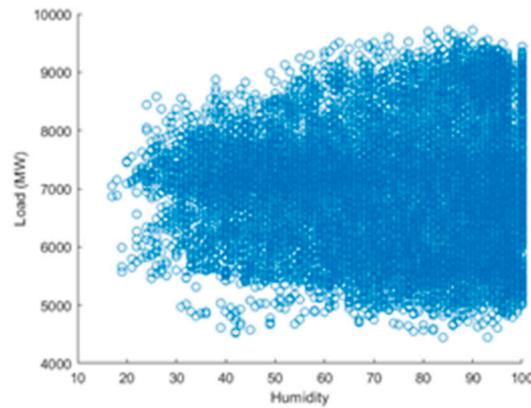


Figure 4. Energy load—humidity influence.

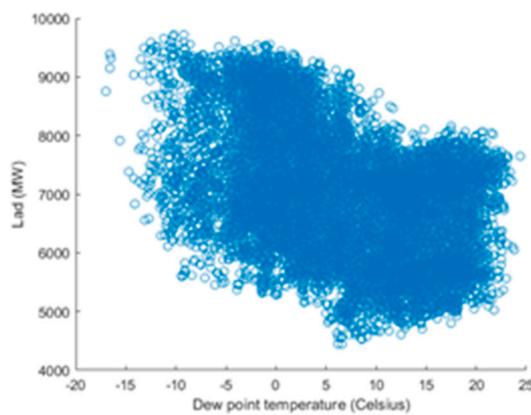


Figure 5. Energy load—dew point influence.

3. Empirical Wavelet Transform

EWT [30] is used to preprocess the energy load data. This will ensure that the analyzed data will follow a smooth path that shapes accordingly to data changes without capturing the peak values introduced by field equipment. All data are preprocessed before using the forecasting algorithm. This method is not used in real time because data are collected until the end of the day. After acquiring the new data set, the EWT preprocess starts and the result is then added to the historical data. This wavelet method adapts to the processed signal by building a set of bandpass filters. To construct the wavelet filter bank, a robust peak detection is used prior to determining a maxima for the spectrum segmentation. This method is proposed based on multiple studies that have validated it as a perfect fit for non-stationary time signals as an energy load measurements vector [31].

To perform a short description of the algorithm, first, a limit, ω_n , is defined between each step segment. The segments are denoted with $A_n = [\omega_{n-1}, \omega_n]$, and because $\omega_n \in (0, \pi)$, the $\cup_{n=1}^N A_n = [0, \pi]$, where N is the number of segments. The scaling function is defined below:

$$\hat{\phi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \leq (1 - \gamma)\omega_n \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\gamma\omega_n} (|\omega| - \omega_n + \gamma\omega_n) \right) \right] & \text{if } (1 - \gamma)\omega_n \leq |\omega| \leq (1 + \gamma)\omega_n \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where function $\beta(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x \geq 1 \end{cases}$ and $\beta(x) + \beta(1 - x) = 1, \forall x \in [0, 1]$ and $\gamma \in (0, 1)$.

The corresponding wavelets are then defined as:

$$\hat{\psi}_n(\omega)(2) = \begin{cases} 1 & \text{if } (1 + \gamma)\omega_n \leq |\omega| \leq (1 - \gamma)\omega_{n+1} \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\gamma\omega_{n+1}} (|\omega| - \omega_{n+1} + \gamma\omega_{n+1}) \right) \right] & \text{if } (1 - \gamma)\omega_{n+1} \leq |\omega| \leq (1 + \gamma)\omega_{n+1} \\ \sin \left[\frac{\pi}{2} \beta \left(\frac{1}{2\gamma\omega_n} (|\omega| - \omega_n + \gamma\omega_n) \right) \right] & \text{if } (1 - \gamma)\omega_n \leq |\omega| \leq (1 + \gamma)\omega_n \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

After defining the wavelet methodology, the detailed coefficients and the approximation coefficients required for the signal reconstruction must be defined. The detailed coefficients are obtained from the inner product of the signal with the empirical wavelets, and the approximation coefficients result from the inner product of the signal with the scaling function:

$$\mathcal{W}_f^\varepsilon(n, t) = \langle f, \psi_n \rangle = \int f(\gamma\omega) \overline{\psi_n(\gamma\omega - t)} d\omega \quad (3)$$

$$\mathcal{W}_f^\varepsilon(0, t) = \langle f, \phi_1 \rangle = \int f(\gamma\omega) \overline{\phi_1(\gamma\omega - t)} d\omega \quad (4)$$

The signal reconstruction is obtained by summing the approximation with all the detailed coefficients, as shown below:

$$f(t) = \mathcal{W}_f^\varepsilon(0, t) * \phi_1(t) + \sum_{n=1}^N \mathcal{W}_f^\varepsilon(n, t) * \psi_n(t) \quad (5)$$

For the energy load signal, a level three decomposition with the proposed EWT method is applied. A graphical comparison of the original signal and the reconstructed signal after processing is illustrated below in Figure 6. As presented, the EWT, smoothens the input data which will prove very beneficial for the forecasting algorithm because it will help form patterns more easily.

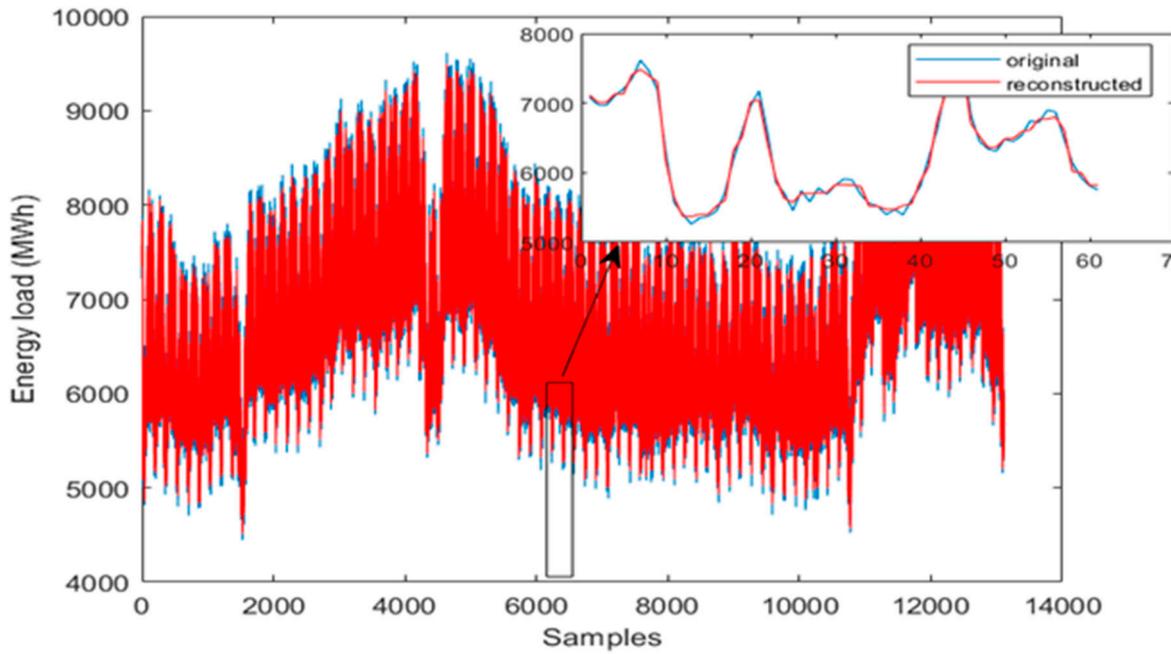


Figure 6. Comparison between reconstructed empirical wavelet transform (EWT) signal and the original signal.

4. Energy Load Forecasting Method

For the proposed energy load prediction, several forecasting models were tested. The best match for the above-presented historical data was an AR model combined with GPR and an error-adjusting function specific to the datasets.

First, a simple AR model is presented, beginning with the mathematical representation [32]:

$$y_m = ct + \sum_{i=1}^n c_i y_{m-1} + e_m \tag{6}$$

where c_i is the method coefficient for the current iteration and represents how the previous value influences the current value, e_m represents the error for the m th value, and ct is the first coefficient of the function with a constant value.

To compute the order of the AR model, a partial autocorrelation function is used with Yule-Walker equations:

$$p_i = \sum_{j=1}^k \Phi_{pj} * \sum_{t=1}^{M-i} (y_m - \bar{y})(y_{m+1} - \bar{y}) \frac{1}{ct * M} \tag{7}$$

where k is the value at which the correlation function Φ_{pk} cuts off after exceeding the model order, M represents the number of observations, and ct is the mean value iteration of the finite time series.

The next step for the method after determining the order of the model is to estimate the model coefficients. To accomplish this, the first equation is expressed in vector form to express the interconnection between the input and the output:

$$\underbrace{\begin{bmatrix} y_m \\ y_{m-1} \\ \vdots \\ y_{m-w+1} \end{bmatrix}}_y = \underbrace{\begin{bmatrix} 1 & y_{m-1} & y_{m-2} & \cdots & y_{m-n} \\ 1 & y_{m-2} & y_{m-3} & \cdots & y_{m-n+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_{m-w} & y_{m-w+1} & \cdots & y_{m-n+w} \end{bmatrix}}_X \underbrace{\begin{bmatrix} c_t \\ c_1 \\ \vdots \\ c_n \end{bmatrix}}_c + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{n+1} \end{bmatrix} \tag{8}$$

where w represents the number of input/output pairs that are constructed using the original time series. Having written the equation in vector form, we can determine the coefficients of the model using the least squares method:

$$C = (X^T X)^{-1} X^T y \tag{9}$$

Next, the proposed GPR model is presented [33]. It can be written as $f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{m}(\mathbf{x}), \mathbf{k}(\mathbf{x}, \mathbf{x}'))$, an expression of its mean and covariance functions. The expression of the output regarding the n -order input vector is computed as follows:

$$y = f(x) + e \tag{10}$$

where e represents the noise, and may be approximated with a Gaussian distribution of zero-mean value, a variance of σ_n^2 , x is the input vector, and y is the scalar output:

$$x_i = [y_{m-i}, y_{m-i+1}, \dots, y_{m-i+n-1}]^T \tag{11}$$

The model, together with the noise assumption, determines the likelihood that represents the probability density based on the parameters. The objective is to construct a training set $D = \{(x_i, y_i) | i = 1, 2, \dots, w\}$ using the input vector and scalar output described above:

$$p(\mathbf{y}|\mathbf{f}) = \mathcal{N}(\mathbf{y}|\mathbf{f}, \sigma_n^2 I) \tag{12}$$

The likelihood presented in the above expression is given by the Gaussian distribution with $\mathbf{y}|\mathbf{f}$ mean and $\sigma_n^2 I$ variance, $\mathbf{y} = [y_1, y_2, \dots, y_w]^T$, $\mathbf{f} = [f(x_1), f(x_2), \dots, f(x_w)]^T$, and I being the unit matrix of w order. Marginal distribution $p(\mathbf{f})$ is defined as having mean zero and a Gram K matrix:

$$p(\mathbf{f}) = \mathcal{N}(\mathbf{f}|\mathbf{0}, K), \quad K_{i,j} = k(x_j, x_i) \tag{13}$$

With the above knowledge, the marginal distribution of y is computed:

$$p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f})d\mathbf{f} = \mathcal{N}(\mathbf{f}|\mathbf{0}, K + \sigma_n^2 I) \tag{14}$$

To ground the predictions in theory, for the predicted output y_* with the input vector x_* , all the parameters' values are averaged and weighted according to their posterior probability:

$$\begin{bmatrix} y_* \\ \mathbf{y}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K_y & \mathbf{k}_* \\ \mathbf{k}_*^T & k_{**} + \sigma_n^2 \end{bmatrix}\right) \tag{15}$$

If there are W training data points, then \mathbf{k}_* is the W order matrix of the covariances evaluated with all the data points for training and testing. Having considered the constraints of the conditioning Gaussians [33], the prediction can be determined:

$$p(y_*|\mathbf{y}) \sim \mathcal{N}(\mathbf{k}_*^T K_y^{-1} \mathbf{y} | k_{**} - \mathbf{k}_*^T K_y^{-1} \mathbf{k}_* + \sigma_n^2) \tag{16}$$

where $k_{**} = k(x_*, x_*)$ and K_y is the covariance matrix.

The kernel, or covariance, function is that which defines the distance and similarity between the training data points. After testing various functions, the best performing kernel function was the RQ:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_n^2 \left(1 + \frac{(\mathbf{x} - \mathbf{x}')^2}{2\alpha l^2} \right)^{-\alpha} \tag{17}$$

This can also be represented as an infinite sum of squared exponential covariance functions with different length scale characteristics (l) and scale mixture (α). RQ kernel was determined as the best choice to minimize the distance between two neighbors through experimental testing. It was decided that for this specific data string from all tested functions the best match was RQ. After integrating the kernel function, the prediction mean and covariance may be calculated:

$$\begin{cases} m(x_*) = \mathbf{c}^T \mathbf{x}_* + \mathbf{k}_*^T K_y^{-1} (\mathbf{y} - \mathbf{m}) \\ \sigma^2(x_*) = k_{**} - \mathbf{k}_*^T K_y^{-1} \mathbf{k}_* + \sigma_n^2 \end{cases} \tag{18}$$

This paper's forecasting method is based on a hybrid of the autoregressive model and a double GPR method, which takes a predicted error into consideration:

$$y = h(x) + \mathbf{c}^T \mathbf{x} + \epsilon(x) \tag{19}$$

where, $h(x) \sim \mathcal{GP}(0, k(x, x))$, c represents the method coefficients matrix and $\epsilon(x)$ represents the predicted residuals. This formulation indicates that the output is the summation of a zero-mean Gaussian process, a set of fixed basis functions, and the zero-mean process of the residuals.

5. Results and Analysis

To test the method’s accuracy, three of the most commonly used performance metrics in machine learning were used: mean absolute percentage error (MAPE), root mean square error (RMSE) and mean absolute error (MAE) [34]. Each of these metrics measures the difference between the experimental and the predicted data. The smaller their value, the better the algorithm will perform:

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{y_i - y_{*i}}{y_i} \right|}{n} * 100\% \tag{20}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - y_{*i})^2}{n}} \tag{21}$$

$$MAE = \frac{\sum_{i=1}^n |y_i - y_{*i}|}{n} \tag{22}$$

where y_i and y_{*i} are the measured and predicted parameters at i time and n represents the number of the considered data.

The algorithm was trained and tested using the energy load historical data described above. Before starting the validation process, and in order to enhance the process some measurements against overfitting are required. Overfitting represents the set of data that corresponds too closely or exactly to another particular set of data and may therefore fail to predict future observations reliably. To protect against overfitting, a fivefold cross-validation method was used. The described performance metrics of the algorithm are MAPE= 4.89%, RMSE= 245.3 MWh, MAE=188.4 MWh. Figure 7 presents a comparison between the predicted and actual values of the load. The first observation is that the forecasted load mimics the characteristics of the historical data. Also, as computed above, the results are expected to generate solid forecasting data for this specific energy load trend.

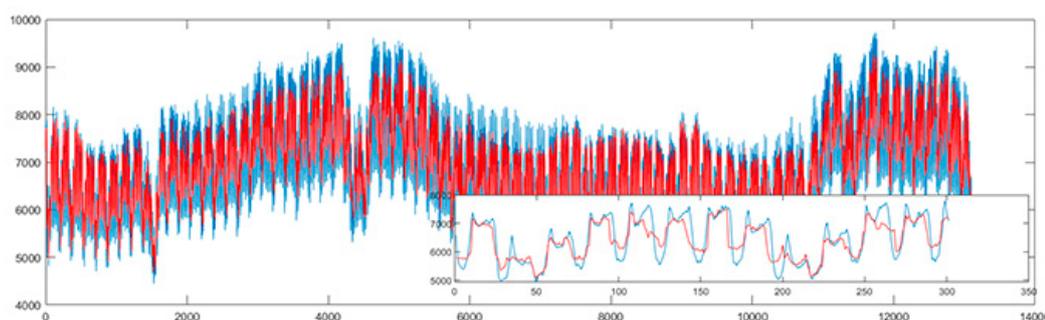


Figure 7. Prediction comparison.

The graphical representations from Figure 8 present the residuals with true and predicted responses, and the influences of temperature and month. It can be observed how the month of the year influences the residual precision, with the lowest values for the middle of the year. The same patterns can be observed for the temperature predictor; in this case, the residuals are higher for lower values of the temperature.

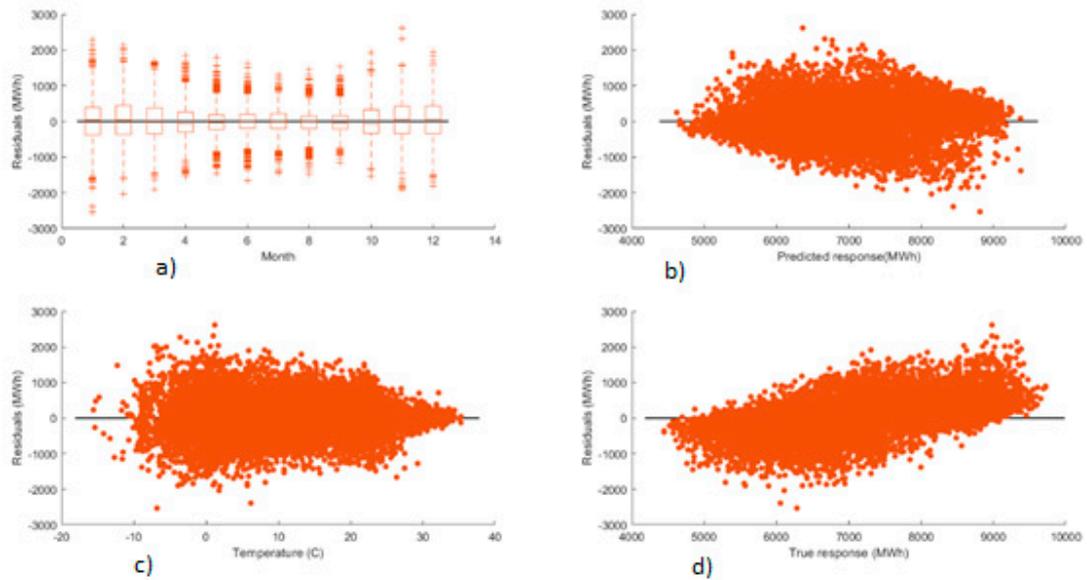


Figure 8. Residuals plot: a) month predictor; b) predicted response comparison; c) temperature predictor; d) true response comparison.

As proposed in the methodology section, to further improve the algorithm, the residuals were predicted using the same regression methodology. Applying this dual-layer prediction process improves the forecasting values and the performance metrics: MAPE= 4.09%, RMSE=221.6 MWh, and MAE=153.2 MWh. Figure 9 illustrates the load’s actual value in blue, the predicted value in red, and the dual-layer prediction in magenta.

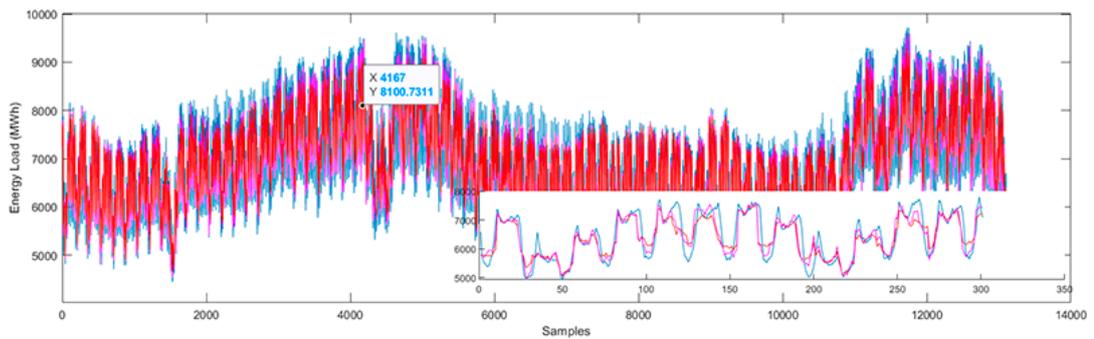


Figure 9. Comparison between historical data, predicted value, and forecasting value with error prediction.

To further validate and test the algorithm’s fitness for the specific load pattern, another test was performed. As presented above, the algorithm was implemented on historic energy load data from January 2018 to June 2019. Data collected from the month of July were used to test the algorithm’s performance. Figure 10 presents a graphical representation of the measured and predicted energy load values for this specific period of time. This graph validates the algorithm that uses residual prediction and also shows its superiority over the hybrid method without error prediction.

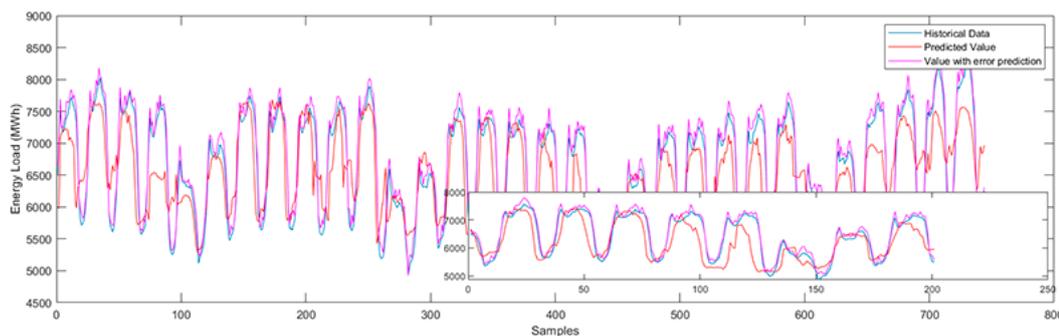


Figure 10. Validation plot for July load forecasting.

A comparison was made between the proposed forecasting method with and without error prediction. Table 1 presents the performance metrics for the July testing.

Table 1. Performance metrics for July testing.

Performance metrics	Simple forecasting	With error prediction
MAPE (%)	5.04	4.28
RMSE (MWh)	205.4	158.7
MAE (MWh)	188.7	138.5

6. Conclusions

Accurate electricity load forecast plays an essential role in planning for utilities and electricity markets in electricity power networks. This paper proposes a novel hybrid GPR forecasting method of the aggregated energy load of a mixture of residential, commercial, and industrial clients. The relatively new EWT method was used for data preprocessing. To ensure a precise load prediction, some of the most important influencing factors were selected. Among these, the most influential were temperature, working/non-working day, and humidity. The hybrid method assumes the combination of an autoregressive time series model and GRP. Using the same methodology, further forecasting of the residuals was added. Following implementation of the algorithm, the method was validated on a new set of data. It was observed that a better approach would be to further apply an error prediction for the residuals. The proposed forecasting method was evaluated using three performance metrics. For the presented load data, this algorithm presents good results that mimic the characteristics of the aggregated plot. The novelty of the approach consists of applying EWT to preprocess the data, combining an AR method with GPR, and using the generated residuals as predictors to increase the algorithm's precision. A future development of the method may be developed by adding a closed-loop system methodology in which the method is executed until the residuals decrease no more.

Author Contributions: All the authors contributed to this study. C.D., H.G.B, D.D.M.: conceptualization, investigation and writing of the original draft; A.T., S.P., I.B.: funding acquisition and project administration; A.T., S.U.: software and data curation; C.D., H.B.: resources; D.M., S.P., S.D.C.: supervision and writing of review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI – UEFISCDI, project number PN-III-P1-1.2-PCCDI-2017-0404 / 31PCCDI/2018, within PNCDI III.

Conflicts of Interest: The authors declare no conflicts of interest.

Nomenclature

RES	Renewable Energy Sources
DR	Demand response
AR	Autoregressive
GPR	Gaussian process regression
EWT	Empiric wavelet transform

RQ	Rational quadratic
MAPE	Mean absolute percentage error
RMSE	Root mean square error
MAE	Mean absolute error

References

1. T. I. E. Agency, Global Energy & CO2 Status Report. 2017. Available online: <https://www.iea.org/> (accessed on 1 May 2020).
2. Jiang, P.; Liu, F.; Song, Y. A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short term load forecasting. *Energy* **2017**, *119*, 694–709, doi:10.1016/j.energy.2016.11.034.
3. Ahmad, T.; Chen, H. Utility companies strategy for short-term energy demand forecasting using machine learning based models. *Sustain. Cities Soc.* **2018**, *39*, 401–417, doi:10.1016/j.scs.2018.03.002.
4. Chen, C.; Wang, F.; Zhou, B.; Chan, K.W.; Cao, Y.; Tan, Y. An interval optimization based day-ahead scheduling scheme for renewable energy management in smart distribution systems. *Energy Convers. Manag.* **2015**, *106*, 584–596, doi:10.1016/j.enconman.2015.10.014.
5. Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, doi:10.1016/j.enconman.2019.111799.
6. Wang, H.; Yi, H.; Peng, J.; Wang, G.; Liu, Y.; Jiang, H.; Liu, W. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy Convers. Manag.* **2017**, *153*, 409–422, doi:10.1016/j.enconman.2017.10.008.
7. Pinto, R.; Bessa, R.J.; Matos, M.A. Multi-period flexibility forecast for low voltage prosumers. *Energy* **2017**, *141*, 2251–2263, doi:10.1016/j.energy.2017.11.142.
8. Mak, S.T. *Advanced metering and demand response applications in New Technologies for Smart Grid Operation*; IOP Publishing: Chicago, USA, 2015; doi: 10.1088/978-0-7503-1158. ISSN 978-0-7503-1158-8.
9. Diamantoulakis, P.D.; Kapinas, V.M.; Karagiannidis, G.K. Big Data Analytics for Dynamic Energy Management in Smart Grids. *Big Data Res.* **2015**, *2*, 94–101, doi:10.1016/j.bdr.2015.03.003.
10. Rasul Enayatifar, Hossein Javedani Sadaei, Abdul Hanan Abdullah, Abdullah Gani. Imperialist competitive algorithm combined with refined high-order weighted fuzzy time series (RHWFTS-ICA) for short term load forecasting. *Energy Convers. Manag.* **2013**, *76*, 1104–1116, doi:10.1016/j.enconman.2013.08.039.
11. He, Y.; Zheng, Y. Short-term power load probability density forecasting based on Yeo-Johnson transformation quantile regression and Gaussian kernel function. *Energy* **2018**, *154*, 143–156, doi:10.1016/j.energy.2018.04.072.
12. Hisashi Takeda, Yoshiyasu Tamura, Seisho Sato. Using the ensemble Kalman filter for electricity load forecasting and analysis. *Energy* **2016**, *104*, doi:10.1016/j.energy.2016.03.070.
13. Lynch, C.; O'Mahony, M.J.; Guinee, R.A. Accurate day ahead temperature prediction using a 24 hour Kalman filter estimator. In Proceedings of the 2015 11th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME) Glasgow, UK, 29 June–2 July 2015.
14. Eseye, A.T.; Zhang, J.; Zheng, D. Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. *Renew. Energy* **2018**, *118*, 357–367, doi:10.1016/j.renene.2017.11.011.
15. Wang, Q.; Li, S.; Li, R.; Ma, M. Forecasting U.S. shale gas monthly production using a hybrid ARIMA and metabolic nonlinear grey model. *Energy* **2018**, *160*, 378–387, doi:10.1016/j.energy.2018.07.047.
16. Beccali, M.; Cellura, M.; Brano, V.L.; Marvuglia, A. Forecasting daily urban electric load profiles using artificial neural networks. *Energy Convers. Manag.* **2004**, *45*, 2879–2900, doi:10.1016/j.enconman.2004.01.006.
17. Mandal, P.; Senjyu, T.; Funabashi, T. Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market. *Energy Convers. Manag.* **2006**, *47*, 2128–2142, doi:10.1016/j.enconman.2005.12.008.
18. Amjady, N.; Keynia, F. Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique. *Energy Convers. Manag.* **2009**, *50*, 2976–2982, doi:10.1016/j.enconman.2009.07.016.
19. Singh, S.N.; Mohapatra, A. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. *Renew. Energy* **2019**, *136*, 758–768, doi:10.1016/j.renene.2019.01.031.

20. Sepasi, S.; Reihani, E.; Howlader, A.M.; Roose, L.R.; Matsuura, M.M. Very short term load forecasting of a distribution system with high PV penetration. *Renew. Energy* **2017**, *106*, 142–148, doi:10.1016/j.renene.2017.01.019.
21. Capizzi, G.; Napoli, C.; Bonanno, F. Innovative Second-Generation Wavelets Construction with Recurrent Neural Networks for Solar Radiation Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 1805–1815.
22. Semero, Y.K.; Zhang, J.; Zheng, D. PV power forecasting using an integrated GA-PSO-ANFIS approach and Gaussian process regression based feature selection strategy. *CSEE J. Power Energy Syst.* **2018**, *4*, 210–218.
23. Yang, A.; Li, W.; Yang, X. Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines. *Knowl. -Based Syst.* **2019**, *163*, 159–173, doi:10.1016/j.knosys.2018.08.027.
24. Kushwaha, V.; Pindoriya, N.M. A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast. *Renew. Energy* **2019**, *140*, 124–139, doi:10.1016/j.renene.2019.03.020.
25. Jiang, H.; Zhang, Y.; Muljadi, E.; Zhang, J.J.; Gao, D.W. A Short-Term and High-Resolution Distribution System Load Forecasting Approach Using Support Vector Regression With Hybrid Parameters Optimization. *IEEE Trans. Smart Grid* **2018**, *9*, 3341–3350.
26. Motepe, S.; Hasan, A.N.; Stopforth, R. Improving Load Forecasting Process for a Power Distribution Network Using Hybrid AI and Deep Learning Algorithms. *IEEE Access* **2019**, *7*, 82584–82598.
27. Zhang, C.; Wei, H.; Zhao, X.; Liu, T.; Zhang, K. A Gaussian process regression based hybrid approach for short-term wind speed prediction. *Energy Convers. Manag.* **2016**, *126*, 1084–1092, doi:10.1016/j.enconman.2016.08.086.
28. Transelectrica. Available online: <http://transelectrica.ro/en/web/tel/despre-noi1> (1 May 2020).
29. R. P. Ltd. Weather_in_the_World. Available online: https://rp5.ru/Weather_archive_in_Bucharest_Filaret (accessed on 1 May 2020).
30. Gilles, J. Empirical wavelet transform. *IEEE Trans. Signal Process.* **2013**, *61*, 3999–4010, doi:10.1109/TSP.2013.2265222.
31. Dedinec, A.; Filiposka, S.; Dedinec, A.; Kocarev, L. Deep belief network based electricity load forecasting: An analysis of Macedonian case. *Energy* **2016**, *115*, 1688–1700, doi:10.1016/j.energy.2016.07.090.
32. Jin, Y. *Multi-Objective Machine Learning*; Springer Science & Business Media: Warsaw, Poland, 2006.
33. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Massachusetts, U.S. ISBN 026218253X. 2006. Available online: www.GaussianProcess.org/gpml (accessed on 30 April 2020).
34. Wang, Y.; Zhang, N.; Tan, Y.; Hong, T.; Kirschen, D.S.; Kang, C. Combining Probabilistic Load Forecasts. *IEEE Trans. Smart Grid* **2019**, *10*, 3664–3674, doi: 10.1109/TSG.2018.2833869.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).