

# A New Hybrid Wavelet-Neural Network Approach for Forecasting Electricity

Heni Boubaker, Ben Amor Souhir and Rezgui Hichem

## Abstract:

This study investigates the performance of a novel neural network technique in the problem of price forecasting. To improve the prediction accuracy using each model's unique features, this research proposes a hybrid approach that combines the  $k$ -factor GARMA process, empirical wavelet transform and the local linear wavelet neural network (LLWNN) methods, to form the GARMA-WLLWNN process. In order to verify the validity of the model and the algorithm, the performance of the proposed model is evaluated using data from Polish electricity markets, and it is compared with the dual generalized long memory  $k$ -factor GARMA-G-GARCH model and the individual WLLWNN. The empirical results demonstrated the proposed hybrid model can achieve a better predicting performance and prove that is the most suitable electricity market forecasting technique.

**Keywords:** Electricity price, Forecasting, Empirical wavelet, Neural network,  $k$ -factor GARMA, G-GARCH

**JEL Classification:** C13 C14 C22 C45 C53

Heni Boubaker, Rabat Business School, BEAR LAB (UIR), Technopolis Rabat-Shore, 11100 Rabat-Salé, Morocco  
Ben Amor Souhir, Institute of High Commercial Studies of Sousse (IHEC), LaREMFiq, 4054 Sousse, Tunisia  
Rezgui Hichem ESC Rennes School of Business, 2 Rue Robert d'Arbrissel, 35065 Rennes, France

\* Correspondence: [heniboubaker@gmail.com](mailto:heniboubaker@gmail.com)

## 1. Introduction

In a modern society, electricity has become a necessary commodity. Our daily lives depend on the consumption of electricity in various forms. Rapid evolution of industrialization in the last century has led to a phenomenal growth of electricity consumption and consequently, the tremendous rise in generation of electrical energy. More precisely, electricity is a key factor of production and economic development, since the electricity is an important complementary to other production factors. Hence, the consumption of electricity will rise due to the economic growth, which requires increased investment to enhance the production capacity to cover the real time demand. As a result, economic growth is a key determinant of electricity demand and any variation in electricity supply may influence several demographic and macroeconomic factors, such as gross domestic product (GPD) growth, population growth, etc. The importance of electricity continues to grow, notably for the expanding health and services sectors, for telecommunications technologies, and for energy industry sectors.

However, the electricity spot prices are characterized by extreme load fluctuations, which cause large and infrequent jumps [Clewlow and Strickland 2000; and Weron et al 2004]. Furthermore, electricity prices exhibit some features such as; non-stationarity, high frequency, multiple seasonality (on annual, weekly, and daily levels), [Escribano et al 2011; Koopman et al 2007; and Knittel and Roberts 2005], hard nonlinearity, high volatility, long memory, high percentage of unusual prices, calendar effect, price spikes and mean reversion. Consequently, these behaviours may affect dramatically the spot prices. Therefore, unlike financial markets, electricity markets are characterized by specific price behaviours seeing that electricity is the unique product that cannot be economically stored [Flynn (2005), and Härdle and Trück (2010)]. In this framework, our research focuses on resolving the problem associated to the complexity of forecasting such prices and resolve the issue of the risk related to the high variability of the electricity prices.

Many approaches have been presented in the literature to forecast electricity price. Due to the unique characteristics of electricity and the uncertainty of market and bidding strategies, electricity price forecasts become more complex than power load forecasting. In stark contrast to past markets, current electricity prices are often accompanied by many characteristics make the prediction of electricity prices very difficult. In the field of electricity price forecasting, broadly two methods; parametric and non-parametric, these two approaches are found to have been applied. In statistical models, autoregressive integrated moving average ARIMA (Contreras et al. (2003); and Heping and Jing (2013)), and generalized autoregressive conditional heteroscedasticity GARCH (Garcia et al. (2005); Ghosh and Kanjilal (2014) and Girish. (2016)) are used extensively.

Nevertheless, these models do not consider the long memory behavior that characterizes the electricity prices, to overcome this limitation Granger and Joyeux (1980) and Hosking (1981) developed the fractional autoregressive moving average model. Baillie et al. (1996) and Bollerslev and Mikkelsen (1996) introduced the fractionally integrated generalized autoregressive conditional heteroscedasticity (FIGARCH) process to model finite persistence in the conditional variance. Many studies have applied these methods for the electricity prices (Koopman et al. (2007), Saâdaoui et al. (2012)). In the spectral domain, these processes have a peak for very low frequencies close to the zero frequency. Hence, it is remarkable that ARFIMA model is not able to model the persistent periodic or cyclical behavior in the time series.

To overcome this insufficiency, Gray et al. (1989) introduced a second generation of the long-memory model termed a generalized (seasonal) long-memory or Gegenbauer autoregressive moving average (GARMA) process, which has been established to estimate both the seasonality and the persistence in the data. On the other hand, in the frequency domain, the spectral density is not necessarily unbounded at the origin, as in the case of the ARFIMA model, but at any frequency  $\lambda$ . In this sense, GARMA process presents a long-memory cyclical behavior at a single frequency  $\lambda$ . Woodward et al. (1998) generalized the GARMA process to the  $k$ -factor GARMA process that allows

the spectral density function to be not just located at a single frequency but presented a number  $k$  of frequencies, identified as the Gegenbauer frequencies or G-frequencies. This model has been applied by numerous authors to reproduce the seasonal persistent patterns (Boubaker and Sghaier (2015), Caporale and Gil-Alana (2014), Caporale et al. (2012), Diongue et al. (2009), Soares and Souza (2006) and Diongue et al. (2004)). Concerning the estimation of the parameter's  $k$ -frequency GARMA process, Gray et al. (1989), Woodward et al. (1998) and Beran (1999) considered the time-domain maximum likelihood method. Whitcher (2004) and Boubaker (2015) suggest estimation method in the wavelet domain founded on the maximal overlap discrete wavelet packet transform (MODWPT).

Further, these statistical methods are often based on linear sequence analysis. So, it is often pointed out that the ability to capture nonlinear behavior and rapid changes in the electricity price process is limited, resulting in poor electricity price forecasting performance. To reproduce these patterns, we suggest hybrid modeling in this study. In the literature, many combination methods have been suggested to avoid the deficiencies related to single models (Yu et al. (2005); Armano et al. (2005); Tseng et al. (2002); Zhang (2003); Taskaya and Casey (2005); Valenzuela et al. (2008); Khashei and Bijari (2010); Tan et al. (2010); Sharkey (2002); Shafie-khah et al. (2011); Jiang et al. (2017); Ben Amor et al. (2018) and Jinliang et al.(2018)). Our approaches adopted can be divided into two categories the non-parametric methods such as the neuronal networks, and the parametric models termed generalized GARCH process.

In the first approach, artificial neural networks (ANN) have been frequently adopted in the electricity market. Wang and Ramsay (1998), Szkuta et al. (1999), Anbazhagan and Kumarappan (2014), Ioannis and Athanasios (2016), Harmanjot et al. (2016); and Jesus et al (2018) adopted the neural networks to model and forecast the dynamics of intra-day prices. Zhang and Benveniste (1992) suggested the wavelet neural networks as an alternative to the conventional NNs (such as feedforward NNs) to reduce the weaknesses related to each method. WNs are one hidden layer networks, which adopt a wavelet as an activation function. The WNs have been effectively used in time series forecasting, (Cao et al. (1995); and Cristea et al. (2000)) and in short-term electricity prices forecasting (Bashir and El-Hawary (2000); Yao et al. (2000); Gao and Tsoukalas (2001); Benaouda et al. (2006); and Ulugammai et al. (2007), Pindoriya et al (2009); and Mashud and Irena (2016)). To preserve the advantage related to the local capacity of the wavelet basis functions while not using many hidden layers, Chen et al. (2004) developed a new type of wavelet neural network termed the local linear wavelet neural network (LLWNN). Therefore, this network needs smaller wavelets for a given problem comparing to the wavelet neural networks. Several researchers for the electricity price forecasting have extensively used the LLWNN model (Pany (2011); Chakravarty et al. (2012); Pany et al. (2013); and Athanassios et al. (2015)).

In the second approach, Boubaker (2015) include the GARCH model, suggested by Engle (1982) and Bollerslev (1986) in the  $k$ -factor GARMA adaptation taking into account time varying volatility. In another research, Boubaker and Boutahar (2011) suggest the  $k$ -factor GARMA-FIGARCH to estimate the long memory behavior in the conditional variance of the exchange rate. Nevertheless, these models are not fully satisfactory in the modelling the volatility of intra-daily financial time series. The main feature of such data is the strong evidence of cyclical patterns in the volatility. For this aim, Bordignon et al. (2007, 2010) suggested a new category of GARCH models characterized by periodic long memory behavior termed the generalized long memory GARCH (G-GARCH). In the literature, Bordignon et al. (2007) and Caporin and Lisi (2010) have applied G-GARCH process to estimate the financial series and Diongue et al. (2009) for modelling the electricity spot price.

This paper focuses on resolving the issues of modeling and forecasting some feature of the electricity prices, notably the existing of the seasonal long memory behavior in the conditional mean and the conditional variance. In this vein, this paper provides three contributions. The main one is to improve the forecasting accuracy of the LLWNN model. This objective is achieved through using the wavelet theory to decompose the historical price, instead of introducing it directly to the Network and assess the effect of different levels of decomposition on forecasting accuracy. This technique can allow

the network to detect the existence of seasonal long memory behavior and thus better estimate the data, this novel network is termed WLLWNN.

The rest of the paper is organized as follows; in the next section, we present a brief review of the literature. Section 2 present the econometric methodology which includes the theoretical concepts of the  $k$ -factor GARMA process, the wavelet local linear neural network model, illustrate the hybrid  $k$ -factor GARMA-WLLWNN method and the  $k$ -factor GARMA-G-GARCH process which estimated using a wavelet estimation approach. Section 3 deals with the empirical framework, where the proposed hybrid model is applied to log-return of electricity spot price forecasting and its performance is compared with the individual WLLWNN model, model and the generalized long memory  $k$ -factor GARMA-G-GARCH model, and section 4 wrap up the conclusion.

## 2. Methodology

### 2.1 The GARMA model

The  $k$ -frequency GARMA model generalizes the ARFIMA model, allowing periodic or quasi-periodic movement in the data. Gray et al. (1989) proposed this model. The multiple frequency GARMA model is defined as follows

$$\Phi(L) \prod_{i=1}^k (I - 2v_{m,i}L + L^2)^{d_{m,i}} (y_t - \mu) = \Theta(L)\varepsilon_t \quad (1)$$

Where  $\Phi(L)$  and  $\Theta(L)$  are the polynomials of the delay operator  $L$  such that all the roots of  $\Phi(z)$  and  $\Theta(z)$  lie outside the unit circle. The parameters  $v_{m,i}$  provide information about periodic movement in the conditional mean,  $\varepsilon_t$  is a white noise disturbance sequence with variance  $\sigma_\varepsilon^2$ ,  $k$  is a finite integer,  $|v_{m,i}| < 1$ ,  $i = 1, 2, \dots, k$ ,  $d_{m,i}$  are long memory parameters of the conditional mean indicating how slowly the autocorrelations are damped,  $\mu$  is the mean of the process,  $\lambda_{m,i} = \cos^{-1}(v_{m,i})$ ,  $i = 1, 2, \dots, k$ , denote the Gegenbauer frequencies (G-frequencies). The GARMA model with  $k$ -frequency is stationary when  $|v_{m,i}| < 1$ , and  $d_{m,i} < 1/2$  or when  $|v_{m,i}| = 1$  and  $d_{m,i} < 1/4$ , the model exhibits a long memory when  $d_{m,i} > 0$ . For a GARMA model with a single frequency, when  $v = 1$ , the model is reduced to an ARFIMA( $p, d, q$ ) model, and when  $v = 1$  and  $d = 1/2$ , the process is an ARIMA model. Finally, when  $d = 0$ , we get a stationary ARMA model. Cheung (1993) determines the spectral density function and shows that for  $d > 0$  the spectral density function has a pole at  $\lambda = \cos^{-1}(v)$ , which varies in the interval  $[0, \pi]$ . It is important to note that when  $|v| < 1$ , the spectral density function is bounded at the origin for GARMA processes, and thus does not suffer from many problems associated with ARFIMA models.

### 2.2 The wavelet local linear wavelet neural network (WLLWNN)

In this section, we propose novel neural network-based wavelet decomposition; this model contains into two steps. Firstly, the historical price data has been decomposed into wavelet domain constitutive sub series using wavelet decomposition and introduced, in the second step, into the network (LLWNN) to produce the set of input variables and form the proposed WLLWNN forecasting model. More precisely, in the first step, the historical price data has been decomposed into wavelet domain constitutive sub series using Wavelet Transform, since the electricity price series are highly volatile, corrupted by occasional spikes and follows by multiple seasonality's. Hence, a price series exhibits richer structure and signal-processing techniques, as wavelet Transform is good tools to

bring out the hidden patterns in the prices.

### 2.2.1 Theoretical concepts of wavelet decomposition

Electricity price series exhibits specific features and rich structure; for this reason, signal-processing techniques like Fourier Transfer, Wavelet Transfer are good candidates for bringing out hidden patterns in price series (Nicolaisen et al. (2000)). To tackle the problem related to the specific behavior of electricity price series, wavelets have been adopted since they can produce a good local representation of the signal in both frequency and time domains. Wavelet decomposition is applied for multi-scale analysis of the signal and decomposes the time series signal into one low-frequency sub-series (constitute the approximation part) and some high-frequency sub-series (constitute the detailed part) in the wavelet domain. These constitutive series have improved statistical properties than original price series and consequently, improved forecasting accuracy can be achieved by their suitable utilization. The basic point, using a dilation and translation operations, this technique allows a flexible time-frequency resolution, and can define local features of a given function in a parsimonious way. Wavelets are orthonormal bases attained through dyadically dilating and translating a pair of specially constructed functions denoted by  $\phi$  and  $\psi$ , which are named father wavelet and mother wavelet, respectively, given by

$$\int \phi(t) dt = 1 \text{ and } \int \psi(t) dt = 0. \quad (2)$$

The smooth and the low-frequency part of the time series are detected by means of the father wavelet while the detail and the high-frequency components are defined by the mother wavelet. The obtained wavelet basis is

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k) \text{ and } \psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k). \quad (3)$$

Where  $j = 1, 2, \dots, J$  indexes the scale and  $k = 1, 2, \dots, 2^j$  indexes the translation. The parameter  $j$  is adopted as the dilation parameter of the wave's functions. This parameter  $j$  adjusts the support of  $\psi_{j,k}(t)$  to locally detect the features of high or low frequencies. The parameter  $k$  is employed to relocate the wavelets in the temporal scale. The number of observations limits the maximum number of scales that can be used in the analysis ( $T \geq 2^J$ ). The localization property is a special property of the wavelet expansion, where the coefficient of  $\psi_{j,k}(t)$  reveals information content of the function at approximate location  $k2^{-j}$  and frequency  $2^{-j}$ . By means of wavelets, any function in  $L^2(\mathbb{R})$  can be extended over the wavelet basis, exceptionally, as a linear combination at arbitrary level  $J_0 \in \mathbb{Z}$  through different scales of the type

$$X(t) = \sum_k s_{J_0,k} \phi_{J_0,k}(t) + \sum_{j \geq J} \sum_k d_{j,k} \psi_{j,k}(t) \quad (4)$$

Where  $\phi_{J_0,k}$  a scaling function with the corresponding coarse scale coefficients  $s_{J_0,k}$  and  $d_{j,k}$  are the detail coefficients given respectively by  $s_{J_0,k} = \int X(t) \phi_{J_0,k}(t) dt$  and  $d_{j,k} = \int X(t) \psi_{j,k}(t) dt$ . These coefficients give a measure of the contribution of the corresponding wavelet to the function. The expression (8) denotes the decomposition of  $X(t)$  into orthogonal components at different resolutions and constitutes the wavelet multiresolution analysis (MRA).

In practical applications, we invariably deal with sequences of values indexed by integers rather than functions defined over the entire real axis. Instead of actual wavelets, we use short sequences of values referred to as wavelet filters. The number of values in the sequence is termed the width of the wavelet filter. Hence, the wavelet analysis measured through a filtering perspective is then well suited

to time series analysis. The wavelet coefficients of the discrete wavelet transform, can be considered from the recursive MRA scheme, which is implemented by a two-channel filter bank (i.e. a high-pass wavelet filter  $\{h_l, l = 0, \dots, L-1\}$  and its associated low pass scaling filter  $\{g_l, l = 0, \dots, L-1\}$  satisfying the quadrature mirror relationship given by  $g_l = (-1)^{l+1} h_{L-1-l}$  for  $l = 0, \dots, L-1$ , where  $L \in \mathbb{N}$  is the length of the filter) illustration of the wavelet transform, is divided into decomposition and reconstruction schemes referring to the forward and inverse wavelet transform. Daubechies (1992) defined a useful category of wavelet filters, termed the Daubechies compactly supported wavelet filters and distinguishes between two choices; the extremal phase filters  $D(L)$  and the least asymmetric filters  $La(L)$ .

### 2.2.2 The local linear wavelet neural network (LLWNN)

Chen et al. (2004) proposed a local linear wavelet neural network (LLWNN) for time series forecasting, and they have shown that this model has more accuracy than the traditional WNN. It comprises of input layer, hidden layer and linear output layer. The input data in the input layer of the network are directly transmitted into the wavelet layer. As the hidden layer neurons make use of wavelets as activation functions, these neurons are usually called 'wavelons'. Instead of using multilayered neural networks and its several variants a WLLWNN is used for forecasting the next day and next week electric load in a deregulated environment. According to wavelet transformation theory, wavelets (used as an activation function) in the following form is a family of functions, generated from one single function  $\psi(x)$  by the operation of dilation and translation.

$$\psi(x) = \left\{ \psi_i = |a_i|^{-1/2} \psi\left(\frac{x-b_i}{a_i}\right); \quad a_i, b_i \in \mathbb{R}^n, i \in \mathbb{N} \right\} \quad (5)$$

$$x = (x_1, x_2, \dots, x_n)$$

$$a_i = (a_{i1}, a_{i2}, \dots, a_{in})$$

$$b_i = (b_{i1}, b_{i2}, \dots, b_{in})$$

$\psi(x)$  is localized in both time space and the frequency space, is called a mother wavelet and the parameters  $a_i$  and  $b_i$  are the scale and translation parameters, respectively. Instead of the straightforward weight  $w_i$  (piecewise constant model), a linear model  $v_i = w_{i0} + w_{i1}x_1 + \dots + w_{in}x_n$  is introduced.

The activities of the linear models  $v_i$  ( $i = 1, 2, \dots, n$ ) are determined by the associated locally active wavelet functions  $\psi_i(x)$  ( $i = 1, 2, \dots, n$ ), thus  $v_i$  is only locally significant. Non-linear wavelet basis functions (named wavelets) are localized in both time space and frequency space. Here  $m = n$  and output ( $Y$ ) of the proposed model is calculated as follows

$$Y = \sum_{i=1}^M (w_{i0} + w_{i1}x_1 + \dots + w_{in}x_n) \psi_i(x) \quad (6)$$

The mother wavelet is

$$\psi(x) = \frac{-x^2}{2} e^{\frac{-x^2}{\sigma^2}} \quad (7)$$

$$\psi(x) = e^{-\left(\frac{x-c}{\sigma}\right)^2} \quad (8)$$

$$\text{Where } x = \sqrt{d_1^2 + d_2^2 + \dots + d_n^2} \quad (9)$$



### 2.2.3 Learning Algorithms for optimizing the neural networks

#### The back-propagation algorithm (BP)

The BP is one of the most common learning algorithms in training the neural networks. At the beginning, the parameters are randomly initialized, and then the algorithm measures the error between the output value and the real value, and finally adjusts the weights in the direction of descendent gradient. The learning rate controls the speed of the training process. If this rate is high, the ANN model will learn quicker, but the learning process will never converge if this rate is too high. In contrast, if the learning rate is so low, the ANN model may converge to a local minimum instead of the global minimum. The equations of the BP algorithm are presented in detail in (Burton and Harley (1994)) and they are briefly described below.

The objective function to minimize is given as

$$E = \frac{1}{2} \left[ y_i - \omega_{1,0} \phi_1(x) - \omega_{1,1} p_1 \phi_1(x) - \dots - \omega_{1,p} p_p \phi_1(x) \right] \quad (10)$$

Where  $y_i$  is the desired value,  $\phi(x)$  is the active wavelet functions,  $\omega_{1,0}$  represent the connection weight,  $p$  is the number of input ( $i = 1, \dots, p$ ) and  $l$  is the number of the hidden units ( $j = 1, \dots, l$ ).

The weight is updated from  $i^{th}$  to the  $(i+1)^{th}$  iteration, that is from  $\omega_i$  to  $\omega_{i+1}$  is given by

$$\omega_{i+1} = \omega_i + \Delta \omega_i = \omega_i + \left( r \frac{\partial E_i}{\partial \omega_i} \right), \quad (11)$$

Denote that  $r$  is the learning rate adopted in the LLWNN model, where  $\frac{\partial E}{\partial \omega}$  for all weights are described by the following equations

$$\frac{\partial E}{\partial \omega_{i,0}} = \omega_{i,0} + r * e * \left( \frac{1}{2} \right) * (x_1^2 + x_2^2 + \dots + x_p^2) * \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \dots + (x_p - c_i)^2)) \quad (12)$$

For  $\forall j \neq 0$ ;

$$\frac{\partial E}{\partial \omega_{i,j}} = \omega_{i,j} + r * e * \left( \frac{1}{2} \right) * (x_1^2 + x_2^2 + \dots + x_n^2) * \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \dots + (x_n - c_i)^2)) * x_j \quad (13)$$

That is

$$\frac{\partial E}{\partial \omega_{1,0}} = \omega_{1,0} + r * e * \left( \frac{1}{2} \right) * (x_1^2 + x_2^2 + \dots + x_p^2) * \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \dots + (x_p - c_i)^2)) \quad (14)$$

$$\frac{\partial E}{\partial \omega_{1,2}} = \omega_{1,2} + r * e * \left( \frac{1}{2} \right) * (x_1^2 + x_2^2 + \dots + x_p^2) * \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \dots + (x_p - c_i)^2)) * x_2 \quad (15)$$

$$\frac{\partial E}{\partial \omega_{2,0}} = \omega_{2,0} + r * e * \left( \frac{1}{2} \right) * (x_1^2 + x_2^2 + \dots + x_p^2) * \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \dots + (x_p - c_i)^2)) \quad (16)$$

$$\frac{\partial E}{\partial \omega_{2,1}} = \omega_{2,1} + r * e * \left( \frac{1}{2} \right) * (x_1^2 + x_2^2 + \dots + x_p^2) * \exp(-((x_1 - c_i)^2 + (x_2 - c_i)^2 + \dots + (x_p - c_i)^2)) * x_1 \quad (17)$$

The other weights are also updated in the same way.

### **The particle swarm optimization algorithm (PSO)**

Kennedy and Eberhart (1995) developed the PSO as an optimization technique. In comparison with other learning algorithms, the PSO clearly proved its efficiency. PSO algorithm is established through simulation of bird flocking in two-dimension space. The position of each agent is denoted by  $XY$  axis position and the velocity are represented by  $v_x$  and  $v_y$ . The agent position's adjustment is recognised by the position and the velocity information. The Bird flocking optimizes the objective function. Each agent knows its best value so far ( $pbest$ ) and its  $XY$  position. In addition, each agent knows the best value so far in the group ( $gbest$ ) among ( $pbest$ ). Mainly each agent tries to adjust its position using the following information.

- 1 The distance between current position and  $pbest$ .
- 2 The distance between the current position and  $gbest$ .

Velocity of each agent can be updated by the following equation

$$v_i^{p+1} = wv_i^p + c_1 rand_1 \times (pbest_i - s_i^p) + c_2 rand_2 (gbest - s_i^p) \quad (18)$$

Where  $v_i^p$  is the velocity of agent  $i$  at iteration  $p$ ,  $w$  is the weight function,  $c_j$  is weighting factor,  $s_i^p$  is the current position of agent  $i$  at iteration  $p$ ,  $pbest_i$  is the  $pbest$  of agent  $i$  and  $gbest$  is the  $gbest$  of the group. The velocity, which progressively gets close to  $pbest$  and  $gbest$  can be computed using the above equation. The actual position, which characterises the searching point in the solution space, can be updated using the following equation

$$s_i^{p+1} = s_i^p + v_i^{p+1} \quad (19)$$

The first term of equation (18), denote the previous velocity of the agent. The velocity of the agent is updated through the second and third terms.

The general steps, which describe the optimization of the LLWNN using the PSO algorithm, can be demonstrated as follows:

Step.1 The initial condition is generated for each agent:

The initial searching points ( $s_i^0$ ) and velocity ( $v_i^0$ ) of each agent are habitually generated randomly within the allowable range. Note that the dimension of search space contains all the parameters of the LLWNN (equation 6).

The current searching point is set to  $pbest$  for each agent. The best-evaluated value of  $pbest$  is set to  $gbest$  and the agent number with the best value is stored.

Step.2 The searching points are evaluated for each agent:

The value of the objective function is calculated for each agent. If this calculated value is improved in comparison with the current  $pbest$  of the agent, the  $pbest$  value is replaced by the current value. If the best value of  $pbest$  is better than the current  $gbest$ ,  $gbest$  is replaced by the best value and the agent number corresponding to the best value is stored.

Step.3 Modification of each searching point:

Using equations (18) and (19), the actual searching point of each agent is updated.

Step.4 Verification of the exit condition:

If the number of the current iteration reaches the number of the predetermined maximum iteration, then exit; If else; go to step 2.

Contrary to the BP, the PSO algorithm avoids the convergence to a local minimum, since it is not founded on gradient information (Abbass et al. (2001)). The objective of the PSO is to produce the best set of weights (particle position) where numerous particles are moving to get the best solution, where the total number of weights characterizes the dimension of the search space.



### 2.3 The hybrid $k$ -factor GARMA-WLLWNN model

Our hybrid methodology combines a semi parametric  $k$ -factor GARMA model and the proposed WLLWNN model. The choice of WLLWNN in the hybrid process is inspired by the wavelet decomposition and its local linear modeling ability. Our approach consists into two steps; in the first step, the aim is modeling the conditional mean using a semi parametric  $k$ -factor GARMA model. Conversely, residuals are important in forecasting time series; they may contain some information that is able to improve forecasting performance. Thus, in the second step, the residuals resulting from the first step will be treated according to a novel wavelet local linear wavelet neural network (WLLWNN) model.

Hence, a time series can be written as

$$y_t = \mu_t + \varepsilon_t \quad (20)$$

Where  $\mu_t$  denote the conditional mean of the time series, and  $\varepsilon_t$  is the residuals. Firstly, the main aim is the parametric modelling, therefore the  $k$ -factor GARMA model is used to reproduce the conditional mean (equation 1). Secondly, the residuals from the parametric model are used as a proxy for the corresponding volatility and modeled using the WLLWNN approach.

Let  $\varepsilon_t$  denote the residuals at time  $t$  from the  $k$ -factor GARMA model, then

$$\varepsilon_t = y_t - \hat{\mu}_t \quad (21)$$

Where  $\hat{\mu}_t$  is the forecast value from the estimated relationship (equation 1).

Then, the forecast values and the residuals of the semi-parametric modelling are the results of the first stage. In the second stage, the aim is the modelling of the residuals using the WLLWNN with  $n$  input nodes, the WLLWNN for the residuals is:

$$\varepsilon_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-n}) \quad (22)$$

Where each  $\varepsilon_{t-i}$  is decomposed using the Wavelet Transform (equation 4),  $f$  is a non-linear, non-parametric function determined by the neural network with the reference to the current state of the data, during the training of the neural network. The output layer of the network (equation 6) gives the forecasting results;

$$\hat{y}_t = \hat{\mu}_t + \hat{\varepsilon}_t \quad (23)$$

Hence, this global prediction, represent the result of forecasting both, the conditional mean and the conditional variance of the time series.

### 2.4 The $k$ -factor GARMA-G-GARCH model

The  $k$ -frequency GARMA model assumes that the conditional variance is constant over time. In the empirical studies, it is well recognized that many time series often exhibit volatility clustering, where time series exhibit both high and low periods of volatility. To reproduce these patterns, we extended the  $k$ -factor GARMA model described above by inserting a fractional filter in the conditional variance equation. For this reason, we propose the dual generalized  $k$ -factor GARMA-G-GARCH model that can capture seasonality and long memory dependence in both the conditional mean and the conditional variance.

The fundamental idea of this model is to include the generalized long-memory process into the equation describing the evolution of conditional variance in a GARCH framework. That is why this new class of models is called Gegenbauer-GARCH (G-GARCH). Thus, we consider the following  $k$ -factor GARMA process with G-GARCH type innovations to consider the presence of a time varying conditional variance

$$y_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \quad (24)$$

Where  $\mu_t$  is the conditional mean of  $y_t$  modelling using the following  $k$ -factor GARMA model:

$$\Phi(L) \prod_{i=1}^k (I - 2\nu_{m,i}L + L^2)^{d_{m,i}} (y_t - \mu) = \Theta(L)\varepsilon_t \quad (25)$$

$$\varepsilon_t / I_{t-1} \square \mathcal{D}(0, \sigma_t^2) \quad (26)$$

Where  $\sigma_t^2$  is the conditional variance,  $I_{t-1}$  being the information up to time  $t-1$ ,  $z_t$  is a *i.i.d* random variable with zero mean and unitary variance and  $\mathcal{D}(\square)$  is a probability density function.

To specify the dynamics of the conditional variance, the starting point is the dynamics of  $\varepsilon_t^2$ . We assume that  $\varepsilon_t^2$  follow a  $k$ -factor GARMA model, which describes a cyclical pattern of length  $S$

$$\left[ (I - L)^{d_{v,0}} (I + L)^{d_{v,k}I(E)} \prod_{i=1}^{k-1} (I - 2\nu_{v,i}L + L^2)^{d_{v,i}} \right] \psi(L)\varepsilon_t^2 = \gamma + [I - \beta(L)]\mathcal{G}_t \quad (27)$$

$$P_v(L)\psi(L)\varepsilon_t^2 = \gamma + [I - \beta(L)]\mathcal{G}_t \quad (28)$$

Where  $\psi(L) = 1 - \sum_{i=1}^q \psi_i L^i$  and  $\beta(L) = 1 - \sum_{i=1}^p \beta_i L^i$  are suitable polynomials in the lag operator  $L$

and  $\mathcal{G}_t = \varepsilon_t^2 - \sigma_t^2$  is a martingale difference,  $d_{v,0} = d_v / 2$ ,  $I(E) = 1$  if  $S$  is even and zero otherwise.

With this assumption, the corresponding GARCH-type dynamics for conditional variance is given by;

$$\sigma_t^2 = \gamma + \beta(L)\sigma_t^2 + \left\{ I - \beta(L) - \left[ (I - L)^{d_{v,0}} (I + L)^{d_{v,k}I(E)} \prod_{i=1}^{k-1} (I - 2\nu_{v,i}L + L^2)^{d_{v,i}} \right] \psi(L) \right\} \varepsilon_t^2 \quad (29)$$

This implies that in the G-GARCH framework each frequency has been modelled by means of a specific long-memory parameter  $d_{v,i}$  (differencing parameter of the conditional variance).

When  $d_{v,0} = d_{v,1} = \dots = d_{v,k}$ , all the involved frequencies have the same degree of memory. Model (29) may provide, cases, most of the existing GARCH models. For example, standard GARCH models (included seasonal GARCH (Bollerslev and Hodrick 1992) can be obtained by putting  $d_{v,i} = 0$ ,  $i = 0, 1, \dots, k$ . Similarly, the FIGARCH model is equivalent to  $S = 1$  and  $0 < d_{v,0} < 1$ . It is interesting to mention that generalized long-memory filters, in principle, may be applied to any category of GARCH structure. Nonetheless, due to the constraints needed for conditional variance positivity, G-GARCH models are not always feasible, for this reason, Bordignon et al. (2007) proposed to model the logarithm of the conditional variances. Therefore, a practical computing solution is to apply the filter to a generalized log-GARCH model. This means beginning from the expression

$$P_v(L)\psi(L)[\ln(\varepsilon_t^2) - \tau] = \gamma + [I - \beta(L)]\mathcal{G}_t \quad (30)$$

Where  $P_v(L)$  is the generalized long memory filter introduced into a GARCH structure,  $\mathcal{G}_t = \ln(\varepsilon_t^2) - \tau - \ln(\sigma_t^2)$  is a martingale difference and  $\tau = E[\ln(z_t^2)]$ . The expected  $\tau$  value depends on the distribution of the idiosyncratic shock and ensures that  $\mathcal{G}_t$  is a martingale difference, given that  $\ln(\varepsilon_t^2) = \ln(\sigma_t^2) + \ln(z_t^2)$ . Under the Gaussian assumption  $\tau = -1.27$ . The expression for conditional variance implied by (33) is

$$\ln(\sigma_t^2) = \gamma + \beta(L)\ln(\sigma_t^2) + [I - \beta(L) - P_v(L)\psi(L)][\ln(\varepsilon_t^2) - \tau] \quad (31)$$

Since we are modelling  $\ln(\sigma_t^2)$  instead of  $\sigma_t^2$ , no constraints for variance positivity are necessary. A further approach of bypassing the problem of parameter constraints is to adopt EGARCH versions of our model.

To sum up, the aim of this study consists in modelling the different patterns in the electricity time series to provide the best forecasting methods. For this purpose, we exploit a hybrid methodology based in combining the semi parametric  $k$ -factor GARMA model with a novel neural network named the WLLWNN model. The performances of the proposed hybrid  $k$ -factor GARMA-WLLWNN model is evaluated using data from Polish Electricity markets and compared with the dual generalized long memory  $k$ -factor GARMA-G-GARCH model and the individual WLLWNN, in order to prove the robustness of our proposed hybrid model.

### 3. Empirical Methodology

#### 3.1 Data description

The data considered are hourly spot prices on the Polish electricity market, covering the period between 1<sup>st</sup> of June 2017 and 31<sup>st</sup> of December 2017, in total  $N = 5137$  hourly observations, illustrated in Figure 1. This series are obtained from the official website of Polish Power Exchange market. In this study, we consider data in the first difference logarithm to makes the series stationary and allows to model returns series ( $R_t = \Delta \text{Log} P_t$ ). Therefore, we analyze the log-return electricity spot price series, to study their statistical and econometric features.

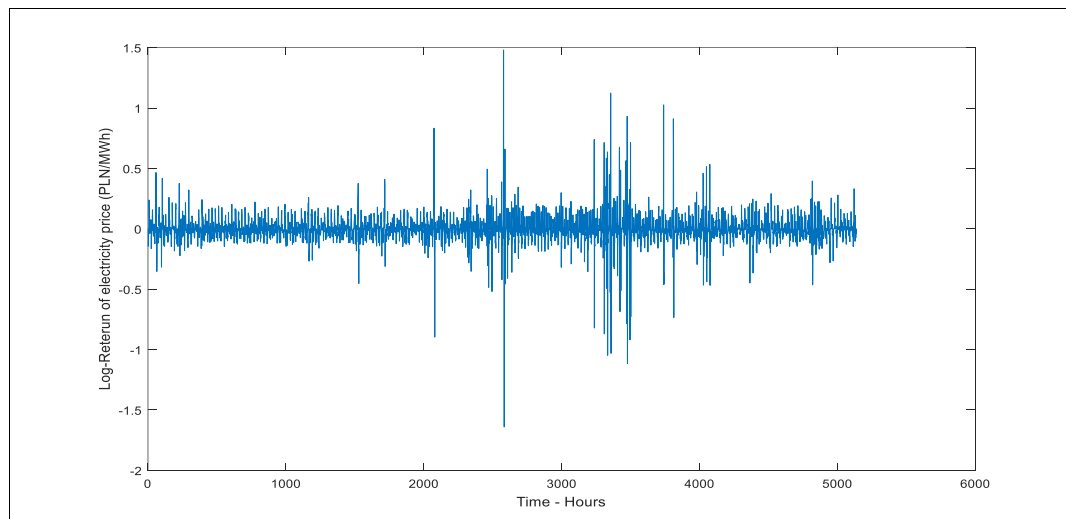


Figure 1. Polish electricity log returns

Figure 1 indicates that the log returns series is stationary. This hypothesis can also be confirmed by the unit root tests (ADF, PP and KPSS). Also, this figure suggests the presence of clustering volatility indicating ARCH effect in the series.

Table 1. Descriptive statistics of the spot prices time series (log-returns)

The log-returns Electricity price	
Mean	$-1.5573 \times 10^{-5}$
Standard deviation	0.1093
Skewness	-0.2814
Kurtosis	36.2357
Jarque-Bera	$2.8116 \times 10^5 ***$

Note: levels of significance are indicated between squared brackets. \*\*\* denotes significance at 1% level.

Descriptive statistics of the log returns data are described in Table 1. The standard deviation is quite small, while the estimated measure of Skewness indicating a non-symmetric distribution. Besides, the large value of the kurtosis statistic, suggests that the underlying data are leptokurtic. This significant departure from normality is also confirmed by the large value of the Jarque-Bera (JB) test. Hence, the electricity spot price series is not normally distributed.

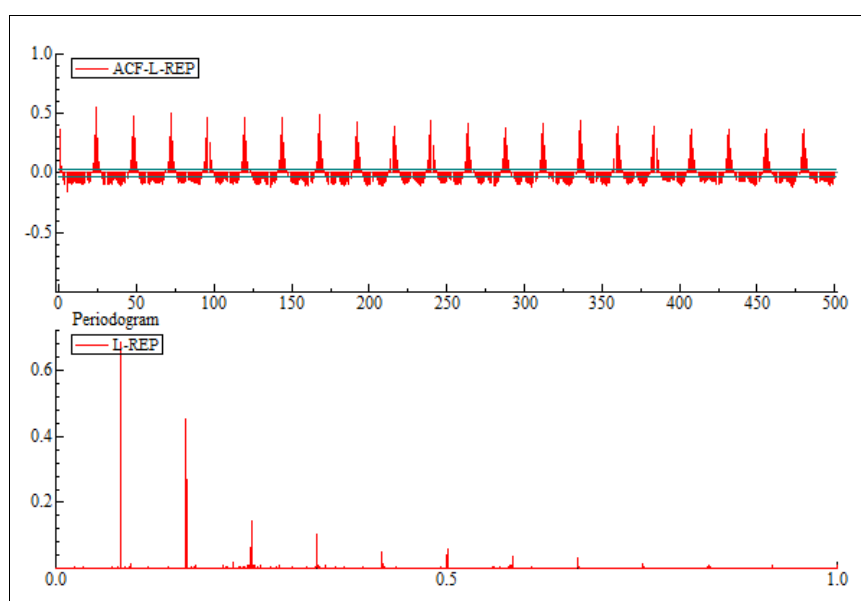
**Table 2. Stationarity test of log-returns series**

	<i>Model (1)</i> <i>With an intercept and a trend</i>	<i>Model (2)</i> <i>With an intercept</i>	<i>Model (3)</i> <i>Without an intercept</i>
<b>ADF</b> <i>t-Statistic</i>	-39.6382 (0.0000)***	-39.6418 (0.0000)***	-39.6457 (0.0000)***
<b>PP</b> <i>t-Statistic</i>	-38.7042 (0.0000)***	-38.8002 (0.0000)***	-38.9822 (0.0000)***
<b>KPSS</b> <i>t-Statistic</i>	0.0008 (0.0000)***	0.0013 (0.0000)***	-

Notes: levels of significance are indicated between squared brackets. \*\*\* denotes significance at 1% level.

We tested for stationary by performing unit root tests, namely, the augmented Dickey-Fuller (ADF, 1976), the Phillips-Perron (PP, 1988) and Kwiatkowski et al. (KPSS, 1992) tests, to the Poland log-returns electricity price. These tests differ in the null hypothesis. The null hypothesis of the ADF and PP tests is that a time series contains a unit root, while the KPSS test has the null hypothesis of stationary. The results of these tests are reported in Table 2; indicating the reject of the hypothesis of non-stationary.

As also illustrated by Figure 2, for the Log-return electricity price series, the spectral density, traced by the periodogram, presents several peaks at equidistant frequencies, which proves the presence of many seasonality's.



**Figure 2. Polish L-REP ACF & Periodogram**

Using the GPH (Geweke and Porter-Hudak, (1983)) and LW (Robinson, (1995)) statistics, we test for the long-range persistence in the conditional mean. Corresponding results shown in Table 3 indicate evidence of long memory.

**Table 3. Results of GPH and LW long-range dependence tests in the conditional mean**

<i>L-REP</i>	<i>Bandwidth</i>	<i>GPH</i>			<i>LW</i>		
		$\hat{d}_m$	<i>Standard error</i>	<i>p-value</i>	$\hat{d}_m$	<i>Standard error</i>	<i>p-value</i>
<i>T=5137</i>	$T^{0.6}=168$	-0.5143	0.0523	0.0000	-0.5458	0.0385	0.0000
	$T^{0.7}=395$	-0.6022	0.0332	0.0000	-0.8150	0.0251	0.0000
	$T^{0.8}=930$	-0.4866	0.0214	0.0000	-0.5894	0.0163	0.0000

### 3.2 Estimation Results

#### ▪ The *k*-factor GARMA estimation results

The seasonality can be easily observed in the frequency domain  $\lambda_i=1/T$ ; where  $\lambda$  is the frequency of the seasonality and  $T$  is the period of seasonality. As shown the spectral densities, represented by periodogram (see Figure 2), are unbounded at equidistant frequencies, which proves presence of multiple seasonality's. They show special peaks at frequencies  $\hat{\lambda}_{m,1}=0.0325$  ( $T=30.76 \approx 30$  hour, 45 min  $\approx 1$  day),  $\hat{\lambda}_{m,2}=0.0904$  ( $T=11.07=11$  hours  $\approx 1/2$  day), and  $\hat{\lambda}_{m,3}=0.1839$  ( $T=5.44=5$  hours, 37 min  $\approx 1/4$  day), corresponding to cycles with daily, semi-daily and quarter-daily periods, respectively.

**Table 4. Estimation of the *k*-factor GARMA model: a wavelet-based approach**

<i>Parameters</i>	<i>k-factor GARMA model estimation</i>
$\hat{\Phi}$	0.5132***
$\hat{\Theta}$	-
$\mu$	-
$\hat{d}_{m,1}$	0.1675***
$\hat{d}_{m,2}$	0.2286***
$\hat{d}_{m,3}$	0.4234***
$\hat{\lambda}_{m,1}$	0.0325***
$\hat{\lambda}_{m,2}$	0.0904***
$\hat{\lambda}_{m,3}$	0.1839***

In second step, we propose to model the conditional variance, so the residuals of the *k*-factor GARMA estimation are shaped through a novel WLLWNN as a first approach and then treated using the generalized GARCH model termed G-GARCH as a second approach, to select the adequate method.

#### ▪ The WLLWNN estimation results

The residuals obtained from the *k*-factor GARMA are considered here as the input of the novel WLLWNN to estimate the conditional variance. For the purpose to avoid the possibility of coupling among different input and to accelerate convergence, all the inputs are normalized within a range of

[0, 1] using the following formula before applying it to the network, which considered as the most commonly used data smoothing method

$$y_{norm} = \frac{y_{org} - y_{min}}{y_{max} - y_{min}} \quad (32)$$

Where  $y_{norm}$  is the normalized value,  $y_{org}$  is the original value,  $y_{min}$  and  $y_{max}$  are the minimum and maximum values of the corresponding residuals data.

#### ▪ Wavelet decomposition

These normalized data are then decomposed using the MODWT with Daubechies least asymmetric ( $La$ ) wavelet filter of length  $L=8$  ( $La(8)$ ). This wavelet filter has been frequently adopted in the financial literature and it has been proved that  $La(8)^1$  provides the best performance for the wavelet time series decomposition. Our MODWT decomposition goes up to level  $J=12$  that is specified by,  $J \leq \log_2 \left[ \frac{T}{L-1} + 1 \right]$  i.e. Where  $T$  represent the length of the given time series and  $L$  denote the length of the filter (Percival and Walden (2000); and Gençay et al. (2002)). The time series is decomposed into 12 details.

#### ▪ The LLWNN modeling

The datasets are presented as follows: (a) A sample of 500 observations to initialize the network training, (b) a training set (4565 observations) and (c) a test set (72 observations). The forecasting experiment is performed over the test set using an iterative forecasting scheme, the model is forecasting for 6, 12, 24, 48 and 72 hours ahead. Details of the datasets are illustrated in the Figure 3.

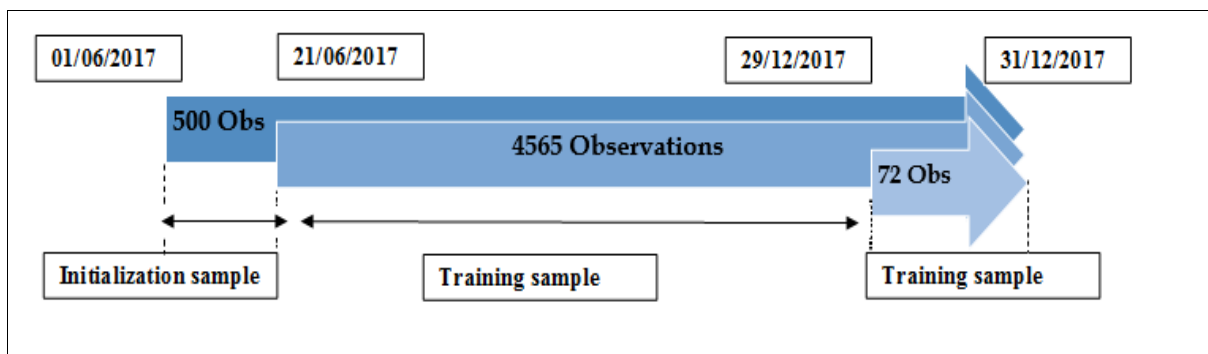


Figure 3. Details of datasets

To find the best neural network architecture, at the beginning the parameters are randomly initialized. Subsequently, using two different algorithms: The Back-Propagation algorithm (BP) and the Particle Swarm Optimization Algorithm (PSO); these parameters are optimized in order to minimize the error between the output values and the real values during the training of the network. Table 5 and Table 6 provide the summary of information related to the network architecture.

<sup>1</sup> As there are no universal selection criteria of the type of wavelets and their width, this choice should be imposed by the objective in order to balance two considerations. Firstly, wavelet filters of too short width can present undesirable artifacts into the multi-resolution analysis. Secondly, the impact of boundary conditions becomes more severe and the localization of MODWT coefficients decreases. For more details concerning this wavelet filter see Daubechies (1992) and Gençay et al. (2002).



**Table 5. LLWNN based BP Algorithm architecture**

<i>Number of hidden layer</i>	10
<i>Learning rate</i>	0.5
<i>Layer conversion function</i>	Wavelet Function
<i>Algorithm</i>	Back Propagation (BP) Learning Algorithm

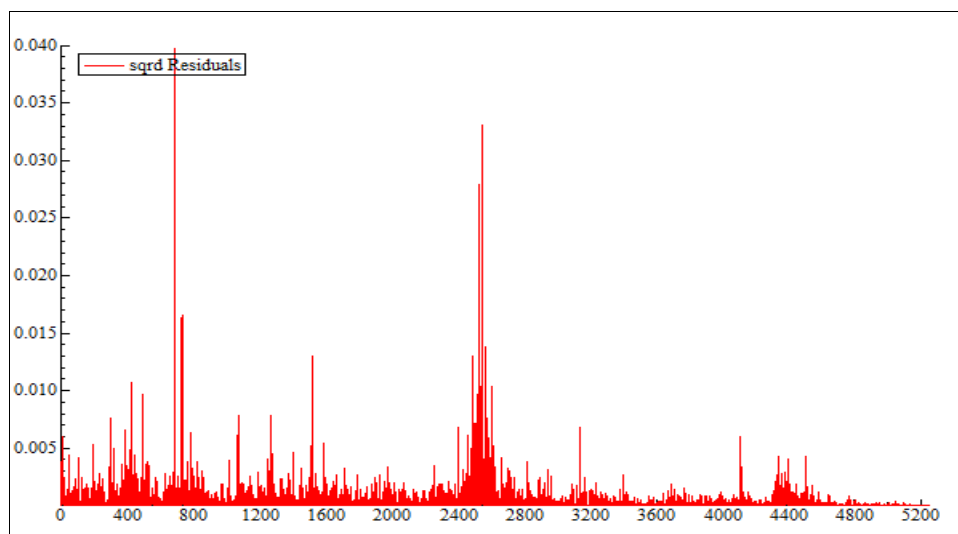
Parameters adopted for running PSO are presented in Table 6.

**Table 6. LLWNN based PSO Algorithm architecture**

<i>Number of populations</i>	20
<i>Number of generations</i>	200
$C_1, C_2$	1.05
<i>Maximum velocity</i>	1
<i>Minimum velocity</i>	0.3
<i>Number of hidden layers</i>	10
<i>Learning rate</i>	0.5
<i>Layer Activation function</i>	Wavelet Function

#### ▪ The $k$ -factor GARMA-G-GARCH estimation results

As shown in Figure 4, for the residuals of the  $k$ -factor GARMA model, the spectral density presents many peaks at equidistant frequencies, which proves the presence of multiple seasonality's. In addition, the squared log-returns are used as a proxy of the corresponding volatility. Long memory tests are performed for the resulted time series. As reported in Table 7, the results of the GPH and LW indicate the presence of long memory in the conditional variance.

**Figure 4. Periodogram of the residuals of the  $k$ -factor GARMA model****Table 7. Results of GPH and LW long-range dependence tests in the conditional variance**

		GPH			LW		
		$\hat{d}_v$	Standard error	$p$ -value	$\hat{d}_v$	Standard error	$p$ -value
$L$ -REP	$T^{0.6}=168$	0.5356	0.0523	0.0000	0.5137	0.0385	0.0000
	$T^{0.7}=395$	0.3566	0.0334	0.0000	0.3923	0.0251	0.0000
	$T^{0.8}=930$	0.2154	0.0216	0.0000	0.2651	0.0251	0.0000

$T=5137$

The residuals from the  $k$ -factor GARMA are modelled based on the G-GARCH process to estimate the seasonal long memory behaviour in the conditional variance. The estimation results of the  $k$ -factor GARMA-G-GARCH model are reported in Table 8

**Table 8. The  $k$ -factor GARMA-G-GARCH Estimation Results**

$k$ -factor GARMA model estimation		The G-GARCH model estimation	
$\hat{\Phi}$	0.6458***	$\hat{\psi}$	0.7250***
$\hat{\Theta}$	-	$\hat{\beta}$	0.5451***
$\mu$	-	$\hat{\gamma}$	-
$\hat{d}_{m,1}$	0.1635***	$\hat{d}_{v,1}$	0.1482***
$\hat{d}_{m,2}$	0.2486***	$\hat{d}_{v,2}$	0.2617***
$\hat{d}_{m,3}$	0.4185***	$\hat{d}_{v,3}$	-
$\hat{\lambda}_{m,1}$	0.0361***	$\hat{\lambda}_{v,1}$	0.0432***
$\hat{\lambda}_{m,2}$	0.0876***	$\hat{\lambda}_{v,2}$	0.1072***
$\hat{\lambda}_{m,3}$	0.1951***	$\hat{\lambda}_{v,3}$	-

The spectral densities, represented by periodogram (Figure 4), are unbounded at equidistant frequencies, which prove presence of several seasonality's. They show special peaks at frequencies  $\hat{\lambda}_{v,1} = 0.0432$  ( $T=23h, 8 \text{ min} \approx 1 \text{ day}$ ), and  $\hat{\lambda}_{v,2} = 0.1072$  ( $T=9h, 20 \text{ min} \approx 1/3 \text{ day}$ ), that corresponding to cycles with daily and third-daily periods, respectively.

### 3.3 Forecasting Results

To evaluate models in a multi-step-ahead forecasting task, we prefer to apply out-of-sample criteria. Therefore, five different periods (6 hours, 12 hours, one day, two days and tree days) were chosen to confirm the quality and the robustness of modeling and forecasting findings. To appraise the forecasting accuracy, two evaluation criteria was applied, the Mean Absolute Error ( $MAE$ ) and the Mean Squared Error ( $MSE$ ), given respectively by:

$$MAE = \frac{1}{N-t_1} \sum_{t=t_1}^N |y_{t+h} - \hat{y}_{t,t+h}| \quad (33)$$

$$MSE = \frac{1}{N-t_1} \sum_{t=t_1}^N (y_{t+h} - \hat{y}_{t,t+h})^2 \quad (34)$$

Where  $N$  is the number of observations,  $N-t_1$  is the number of observations for predictive performance,  $y_{t+h}$  is the log-return series through period  $t+h$ ,  $\hat{y}_{t,t+h}$  is the predictive log-return series of the predictive horizon  $h$  at time  $t$ .

To assess the prediction performance of the proposed hybrid methodology, the  $k$ -factor GARMA-WLLWNN was compared with two models: the individual WLLWNN model and the  $k$ -factor GARMA-G-GARCH model. Concerning the training of the network, we adopted two different learning algorithms (BP and PSO). Furthermore, we adopted five-time horizons; 6 hours, 12 hours, one day, 2 days and 3 days ahead forecasting, using the  $MAE$  and the  $MSE$  out of sample criteria. Further, we also apply the statistical test of Diebold and Mariano (1995) and say that the forecasts have equal accuracy. Under the null hypothesis of no difference, the test statistics noted ( $DM$ ) is asymptotically  $N(0,1)$  distributed. To evaluate the performance of hybrid methodology forecast, we

consider the  $k$ -factor GARMA-G-GARCH model for purpose of comparison between the forecast results of all other models. The forecast findings are given in Table 9.

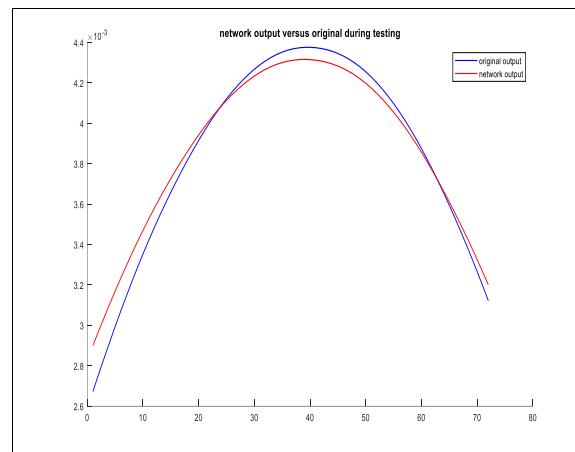
Results reveal that the individual WLLWNN based PSO algorithm outperforms the individual WLLWNN based BP algorithm; this result prove the superiority of the PSO algorithm for training neural network model. This result can be explained by the fact that in the case of the BP algorithms weights are updated in the direction of the negative gradient. Hence, the network training with BP algorithms present some drawbacks such as very slow convergence to a local minimum. Nevertheless, in the case of training with PSO algorithm, weights are characterized by particles position. These particles velocity and position are updated, to search for personal best and global best values. This will avoid the convergence of weights to a local minimum.

Furthermore, we observe from Table 9 that the hybrid  $k$ -factor GARMA-WLLWNN model outperforms all other computing techniques. In fact, this model uses the strength of three techniques at the same time; firstly, the semi-parametric  $k$ -factor GARMA model that allows detecting and estimating both the long memory and the seasonality in the conditional mean. Secondly, the wavelet decomposition, which can produce a good local representation of the signal in both time and frequency domains and hence it's a good tool to bring out the hidden patterns in the electricity prices, such as high volatility, corrupted by occasional spikes and follows by multiple seasonality's. Finally, with the capacity of the LLWNN model as a nonlinear, nonparametric model, and its particularity by having a wavelet activation function and local linearity, this network can capture more subtle aspects of the data. Hence, the proposed hybrid  $k$ -factor GARMA-WLLWNN is a robust tool that can be deal with the features of the electricity prices and provide the best forecasting results.

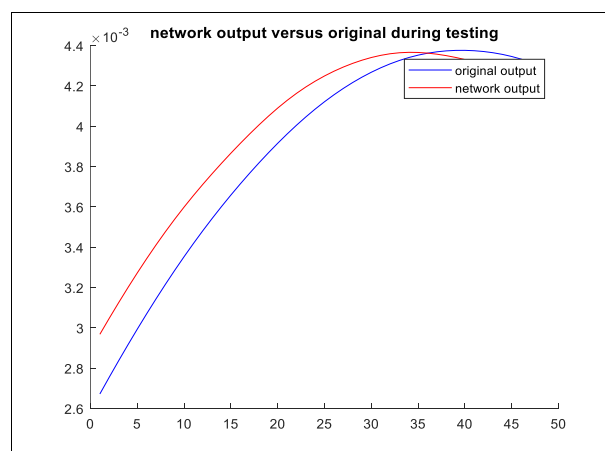
As shown in Figures 5, the predictions of the  $k$ -factor GARMA-WLLWNN model based PSO algorithm for all the five horizons are very close to the real values. That is confirm the forecasting results (Table 9), which indicate that the  $k$ -factor GARMA-WLLWNN process prediction errors are the smallest for all evaluation criteria and for all forecast time horizons.

**Table 9. Out of sample Forecasts Results**

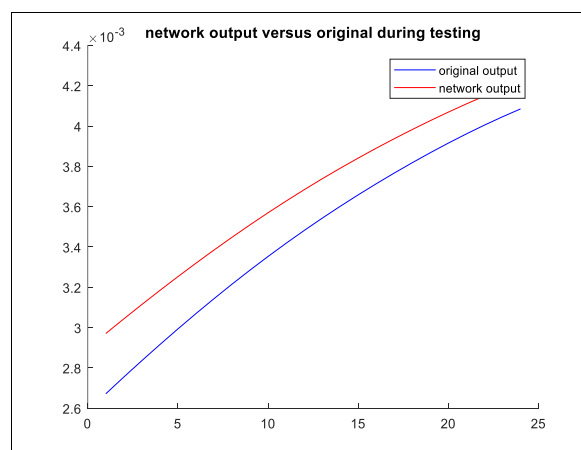
<i>Models</i>	<i>Criterion</i>	$h = 6$	$h = 12$	$h = 24$	$h = 48$	$h = 72$
WLLWNN based BP Algorithm	MAE	$6.3201 \times 10^{-5}$	$1.9792 \times 10^{-5}$	$7.3147 \times 10^{-5}$	$1.0438 \times 10^{-4}$	$2.9156 \times 10^{-4}$
	MSE	$4.0121 \times 10^{-9}$	$2.0457 \times 10^{-10}$	$5.6173 \times 10^{-9}$	$1.2343 \times 10^{-8}$	$7.2376 \times 10^{-8}$
	DM	1.1534	1.2752	1.3684	1.4943	1.6109
WLLWNN based PSO Algorithm	MAE	$2.3212 \times 10^{-7}$	$3.8692 \times 10^{-7}$	$9.9865 \times 10^{-7}$	$2.7763 \times 10^{-7}$	$3.7975 \times 10^{-7}$
	MSE	$6.9811 \times 10^{-14}$	$2.1145 \times 10^{-13}$	$1.1824 \times 10^{-12}$	$1.1458 \times 10^{-13}$	$2.1278 \times 10^{-13}$
	DM	1.3461	1.5272	1.6234	1.9375	2.0154
The hybrid $k$ - factor GARMA- WLLWNN based BP Algorithm	MAE	$4.2367 \times 10^{-5}$	$1.4132 \times 10^{-5}$	$6.3552 \times 10^{-5}$	$7.7246 \times 10^{-5}$	$8.9543 \times 10^{-5}$
	MSE	$1.8561 \times 10^{-9}$	$1.8937 \times 10^{-10}$	$4.8539 \times 10^{-9}$	$7.0336 \times 10^{-9}$	$9.1357 \times 10^{-9}$
	DM	1.9784	2.3256	2.7631	2.9464	3.2546
The hybrid $k$ - factor GARMA- WLLWNN based PSO Algorithm	MAE	$1.4251 \times 10^{-9}$	$1.9542 \times 10^{-8}$	$3.2757 \times 10^{-9}$	$4.2489 \times 10^{-9}$	$2.2146 \times 10^{-8}$
	MSE	$2.0723 \times 10^{-18}$	$5.1127 \times 10^{-16}$	$1.4121 \times 10^{-17}$	$2.7442 \times 10^{-17}$	$7.6058 \times 10^{-16}$
	DM	2.5671	2.9653	3.4378	3.6647	3.8651
The $k$ -factor GARMA-G- GARCH model	MAE	$7.4673 \times 10^{-5}$	$7.8461 \times 10^{-5}$	$3.7852 \times 10^{-7}$	$4.6834 \times 10^{-8}$	$5.1289 \times 10^{-8}$
	MSE	$4.5632 \times 10^{-14}$	$9.5362 \times 10^{-14}$	$1.8457 \times 10^{-15}$	$2.3387 \times 10^{-15}$	$3.7826 \times 10^{-15}$



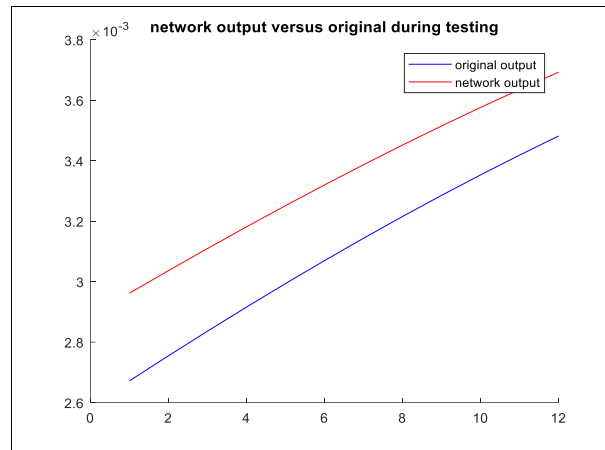
Forecasting with the k-factor GARMA-WLLWNN based PSO h=72 hours



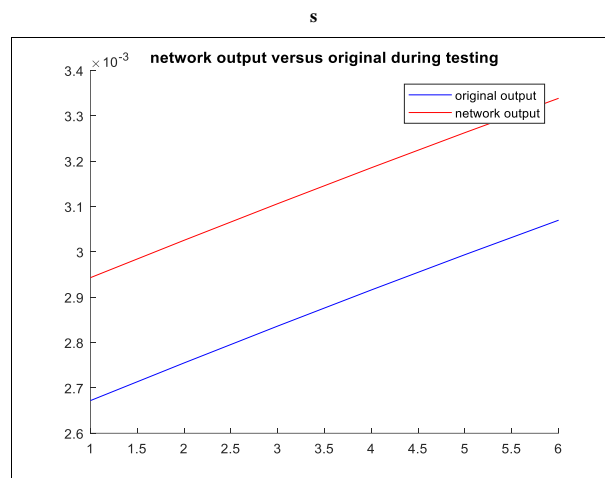
Forecasting with the k-factor GARMA-WLLWNN based PSO h=48 hours



Forecasting with the k-factor GARMA-WLLWNN based PSO h=24 hours



Forecasting with the k-factor GARMA-WLLWNN based PSO h=12 hour



Forecasting with the k-factor GARMA-WLLWNN based PSO h=6 hours

**Figure 5. Forecasting Polish electricity price using the hybrid GARMA-WLLWNN based PSO**

#### 4. Conclusion

In this paper, we develop a combining approach for electricity price forecasting, using jointly the parametric  $k$ -factor GARMA and the novel WLLWNN models. Our new hybrid approach consists into two steps; firstly, the  $k$ -factor GARMA model is adopted to estimate the conditional mean of the time series, since its able to estimate the periodic long memory behavior in the data. Secondly, the residuals from the  $k$ -factor GARMA process are used as a proxy for the corresponding volatility and are estimated based on a novel Neural Network termed the Wavelet Local Linear Neural Network (WLLWNN) model. In this network, the data has been decomposed into wavelet domain constitutive sub series using Wavelet Transform and then introduced into the network to produce the set of input variables and form the proposed WLLWNN forecasting model. On the other hand, when we deal with neural networks it is very important to choose an appropriate algorithm for training, so this paper presents a comparison of two learning algorithms; the BP and PSO algorithms.

The performance of the proposed hybrid model is evaluated using data from Polish Electricity markets. Moreover, it is compared with the dual generalized long memory  $k$ -factor GARMA-GARCH model, and the individual WLLWNN, to prove the robustness of our proposed hybrid model. The empirical results prove that the proposed  $k$ -factor GARMA-WLLWNN method is the most suitable price forecasting technique. Since, it is can produce smaller predicting errors than the other computing techniques. It may consider as a powerful forecasting method, notably when we need higher forecasting accuracy.

## References

1. Abbass, H.A. R. Sarker, and C. Newton. 2001. "A Pareto-frontier Differential Evolution Approach for Multi-objective Optimization Problems". IEEE Congress on Evolutionary Computation. 2: 971-978.
2. Aggarwal. S.K., L.M. Saini, and A. Kumar. 2008. "Electricity Price Forecasting in Ontario Electricity Market Using Wavelet Transform in Artificial Neural Network Based Model". International Journal of Control, Automation, and Systems. 6 (5): 639-650.
3. Anbazhagan. S., Kumarappan. N. 2014. "Day-ahead deregulated electricity market price forecasting using neural network input featured by DCT." Energy Conversion and Management, Volume 78, Pages 711- 719.
4. Armano, G., M. Marchesi, and A. Murru. 2005. "A hybrid genetic-neural architecture for stock indexes forecasting". Information Sciences 170: 3-33.
5. Athanassios M. K., Antonios C., and Pericles A. 2015. "Agent-based Short-Term Load and Price Forecasting Using a Parallel Implementation of an Adaptive PSO-Trained Local Linear Wavelet Neural Network." International Conference on the European Energy Market (EEM).
6. Bashir, Z., and M.E. El-Hawary. 2000. "Short term load forecasting by using wavelet neural networks, in: Electrical and Computer Engineering." IEEE 163-166.
7. Ben Amor, S. Boubaker H., Belkacem L. 2018, " Forecasting electricity spot price for Nord Pool market with a hybrid -factor GARMALLWNN model ", Journal of Forecasting, 37, 832-851.
8. Benaouda, D., G. Murtagh, J.L. Starck, and O. Renaud. 2006. "Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting". Neurocomputing, 70, 139-154.
9. Beran, J. 1999. "SEMIFAR Models-a Semiparametric Framework for Modelling Trends, Long-Range Dependence and Nonstationarity". Center of Finance and Econometrics, University of Konstanz.
10. Bollerslev, T. 1986. "Generalized autoregressive conditional heteroscedasticity". Journal of Econometrics 31: 307-327.
11. Bollerslev, T., and H.O. Mikkelsen. 1996. "Modelling and pricing long memory in stock market volatility. J. Econ. 73 (1): 151-184
12. Bollerslev, T., and R. Hodrick. 1992. "Financial market efficiency tests". The Handbook of Applied Econometrics I. Macroeconomics, North-Holland, Amsterdam.
13. Bordignon, S., M. Caporin, and F. Lisi. 2007. "Generalised long-memory GARCH models for intra-daily volatility". Computational Statistics & Data Analysis 51: 5900-5912.
14. Bordignon, S., M. Caporin, and F. Lisi. 2010. "Periodic long memory GARCH models" Econom Rev 28:60-82.
15. Boubaker, H., and N. Sghaier. 2015. "Semiparametric generalized long-memory modeling of some MENA stock market returns: A wavelet approach". Economic Modelling 50: 254-265.
16. Boubaker. H. 2015. "Wavelet Estimation of Gegenbauer Processes: Simulation and Empirical Application". Comput Econ 46: 551-574.
17. Boubaker; H., and M. Boutahar. 2011. "A wavelet-based approach for modelling exchange rates". Stat Methods Appl 20: 201-220.
18. Bunn, D.W., Karakatsani, N., 2003. "Forecasting electricity prices." London Business School 1.
19. Burton, B., and R.G. Harley. 1994. "Reducing the computational demands of continually online-trained artificial neural networks for system identification and control of fast processes". IEEE IAS Annual Meeting 1836-1843.
20. Cao, L., Y. Hong, H. Fang, and G. He. 1995. "Predicting chaotic time series with wavelet Networks." Physica D 85: 225-238.
21. Caporale, G.M., L.A. Gil-Alana. 2014. "Long-run and cyclical dynamics in the US stock market". J. Forecast 33: 147-161.
22. Caporale, G.M., J. Cuñado, and L.A. Gil-Alana. 2012. "Modelling long-run trends and cycles in financial time series data". J. Time Ser. Anal 34 (2): 405-421.



23. Caporin, M., and F. Lisi. 2010. "Misspecification tests for periodic long memory GARCH models." *Statistical Methods & Applications* 19: 47-62.
24. Chakravarty, S., M. Nayak, and M. Bisoi. 2012. "Particle Swarm Optimization Based Local Linear Wavelet Neural Network for Forecasting Electricity Prices, Energy, Automation, and Signal (ICEAS)." *Energy, Automation, and Signal (ICEAS)*, IEEE: 1-6.
25. Chen, Y., J. Dong, B. Yang, and Y. Zhang, 2004. "A local linear wavelet neural network." *Intelligent Control and Automation*, IEEE: 1954-1957.
26. Cheung, Y.W. 1993. "Long memory in foreign-exchanges rates." *J Bus Econ Stat* 11: 93-101.
27. Clewlow, L., and C. Strickland. 2000. "Energy derivatives pricing and risk management." *Lacima Publications*, London.
28. Contreras, J., R. Espinola, F.J. Nogales, and A.J. Conejo. 2003. "ARIMA models to predict next day electricity prices". *IEEE Transactions on Power systems* 18 (3): 1014-1020.
29. Cristea, P., R. Tuduce, and A. Cristea. 2000. "Time series prediction with wavelet neural networks." *Neural network applications in electrical engineering*.
30. Daubechies, I. 1992. "Ten lectures on wavelets". Philadelphia: SIAM.
31. Diebold, F.X., and Mariano, R. S. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13, 253-263.
32. Diongue, A.K., G. Dominique, and V. Bertrand. 2009. "Forecasting electricity spot market prices with a k-factor GIGARCH process". *Applied Energy* 86: 505-510.
33. Engle, R.F., 1982. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." *Econometrica* 50: 987-1008.
34. Escribano, A., J.I. Peña, and P. Villaplana. 2011. "Modelling electricity prices: international evidence." *Oxford bulletin of economics and statistics* 73: 622-650.
35. Gao, R., and H. I. Tsoukalas. 2001. "Neural-wavelet methodology for load forecasting." *Journal of Intelligent & Robotic Systems* 31: 149-157.
36. Garcia, R.C., J. Contreras., M. Akkeren, and J.B.C. Garcia. 2005. "A GARCH forecasting model to predict day-ahead electricity prices". *IEEE Trans. Power System* 20(2):867-874.
37. Gençay, M., Selçuk, F., & Whitcher, B. 2002. *An introduction to wavelets and other filtering methods in finance and economics*. Waltham: Academic Press.
38. Geweke, J., and S. Porter-Hudak. 1983. "The Estimation and Application of Long Memory Time Series Models." *Journal of Time Series Analysis* 4: 221-238.
39. Girish, G.P. 2016. "Spot electricity price forecasting in Indian electricity market using autoregressive-GARCH models". *Energy Strategy Reviews*, Volumes 11-12, pages 52-57.
40. Ghosh, S., Kanjilal. K. 2014. "Modelling and forecasting of day-ahead electricity price in Indian energy exchange evidence from MSARIMA-EGARCH model", *Int. J. Indian Cult. Bus. Manag.* 8, 413- 423.
41. Granger, C.W.J. 1989. "Combining forecasts: twenty years later". *International Journal of Forecasting*; 8:167-73.
42. Granger, C.W.J., and R. Joyeux. 1980. "An introduction to long-memory time series models and fractional differencing". *J. Time Ser. Anal* 1 (1), 15-29.
43. Gray, H.L., N.F. Zhang, and W.A. Woodward. 1989. "On generalized fractional processes". *J. Time Ser. Anal* 10 (3): 233-257.
44. Harmanjot, S.S. Liping, F., Ling., Guan. 2016. "Forecasting day-ahead price spikes for the Ontario electricity market." *Electric Power Systems Research*, Volume 141, Pages 450- 459.
45. Heping. L and Jing., S .2013. "Applying ARMA-GARCH approaches to forecasting short-term electricity prices". *Energy Economics*, Volume 37, Pages 152-166.
46. Hosking, J.R.M. 1981. "Fractional differencing". *Biometrika* 68 (1): 165-176.

47. Huisman, R., Huurman, C., and Mahieu, R., 2007. "Hourly electricity prices in day-ahead markets." *Energy Economics* 29, 240-248.
48. Ioannis PP., Athanasios S.D. 2016. "Day-ahead electricity price forecasting via the application of artificial neural network based models." *Applied Energy*, Volume 172, Pages 132- 151.
49. Jesus. L., Fjo De. R., Peter. V., Bart De. S. 2018. "Forecasting day-ahead electricity prices in Europe: The importance of considering market integration." *Applied Energy*, 211; 890-903.
50. Jiang P, Liu F, Song YL. 2017. "A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting." *Energy*; 119:694-709.
51. Jinliang. Z., Yi-Ming. W., Dezhi. L., Zhongfu. T., Jianhua. Z. 2018. "Short term electricity load forecasting using a hybrid model." *Energy*.
52. Kennedy, J.R. and Eberhart. 1995. "Particle swarm optimization". *IEEE international conference on neural networks*. 4: 1942-1948.
53. Knittel, C., and M.R. Roberts. 2005. "An empirical examination of restructured electricity prices." *Energy Economics* 27(5): 791-817.
54. Khashei, M. and M. Bijari. 2010. "An artificial neural network model for time series forecasting." *Expert Syst. Appl.* 37: 479-489.
55. Koopman, S.J., M. Ooms, and M.A Carnero. 2007. "Periodic seasonal Reg ARFIMA-GARCH models of daily electricity spot prices." *Journal of the American Statistical Association* 102(477): 16-27.
56. Kwiatkowski, D., P.C. Phillips, P. Schmidt, and Y. Shin. 1992. "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?". *Journal of Econometrics* 54 (Issues 1-3): 159-178. Mallat, S. 1999. "A wavelet tour of signal processing". Academic press.
57. Mallat, S., and Zhang. 1993. "Matching pursuits with time-frequency dictionaries." *IEEE Transactions on Signal Processing* 41: 3397-3415.
58. Mashud. R., Irena. K. 2016. "Forecasting electricity load with advanced wavelet neural networks." *Neurocomputing*, Volume 182, Pages 118-132.
59. Nicolaisen. J. D., C.W. Richter, and G. B. Sheble. 2000. "Price signal analysis for competitive electric generation companies". *Proc. of Conf. Electric Utility Deregulation and Restructuring and Power Technologies*, London, U.K., 4-7.
60. Pindoriya. N. M., Singh. S. N., Singh. S. K. 2009. "An Adaptive Wavelet Neural Network-Based Energy Price Forecasting in Electricity Markets." *IEEE Transactions on Power Systems*, Volume: 23, Issue: 3
61. Pany, P.K. 2011. "Short-Term Load Forecasting using PSO Based Local Linear Wavelet Neural Network." *International Journal of Instrumentation* 1: Issue-2.
62. Pany, P.K., and S.P. Ghoshal. 2013. "Day-ahead Electricity Price Forecasting Using PSO-Based LLWNN Model." *International Journal of Energy Engineering (IJEE)* 3: 99-106.
63. Percival, D.B., Walden, A.T., 2000. *Wavelet Methods for Time Series Analysis*. Cambridge Press, Cambridge.
64. Robinson, P.M. 1995. "Log-Periodogram Regression of Time Series with Long-Range Dependence." *Annals of statistics* 23: 1048-1072.
64. Saâdaoui, F., N. Chaâben, and S. Benammou. 2012. "Modelling power spot prices in deregulated European energy markets: a dual long memory approach." *Global Business and Economics Review* 14: No. 4.
65. Shafie-khah. M., Parsa Moghaddam. M., Sheikh-El-Eslami. M.K. 2011. "Price forecasting of day-ahead electricity markets using a hybrid forecast method". *Energy Conversion and Management* 52, 2165- 2169.
66. Sharkey, A.J., 2002. "Types of multinet system". *International Workshop on Multiple Classifier Systems*. Springer: 108-117.
67. Soares, L.J., and L.R. Souza. 2006. "Forecasting electricity demand using generalized long memory." *International Journal of Forecasting* 22, 17-28.

68. Szkuta, B., L. Sanabria, and T. Dillon. 1999. "Electricity price short-term forecasting using artificial neural networks." *IEEE Trans Power Syst* 14: 851-857.
69. Tan, Z., J. Zhang, J. Wang, and J. Xu. 2010. "Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models." *Appl Energy* 87:3606-3610.
70. Taskaya, T., and M.C. Casey. 2005. "A comparative study of autoregressive neural network hybrids." *Neural Networks* 18: 781-789.
71. Tseng, F.M., H.C. Yu, and G.H. Tzeng. 2002. "Combining neural network model with seasonal time series ARIMA model." *Technological Forecasting & Social Change* 69: 71-87.
72. Ulugammai, M., P. Venkatesh, P.S. Kannan, and N.P. Padhy. 2007. "Application of bacterial foraging technique trained artificial and wavelet neural networks in load forecasting." *Neurocomputing* 70: 2659-2667.
73. Valenzuela, O., I. Rojas, F. Rojas, H. Pomares, L. Herrera, A. Guillen, L. Marquez, and M. Pasadas. 2008. "Hybridization of intelligent techniques and ARIMA models for time series prediction." *Fuzzy Sets Syst.* 159: 821-845.
74. Wang, A., and B. Ramsay. 1998. "A neural network-based estimator for electricity spot-pricing with particular reference to weekend and public holidays." *Neurocomputing* 23: 47-57.
75. Weron, R., I. Simonsen, and P. Wilman. 2004. "Modelling highly volatile and seasonal markets: evidence from the Nord Pool electricity market." *The Application of Econophysics*, Springer, Tokyo 182-191.
76. Weron, R. 2006. "Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach." *The Wiley Finance Series (Book 396)*, Wiley, Chichester.
77. Whitcher, B. 2004. "Wavelet-based estimation for seasonal long-memory processes." *Technometrics* 46.
78. Woodward, W.A., Q.C. Cheng, and H.L. Gray. 1998. "A-factor GARMA long-memory model". *J. Time Ser. Anal.* 19 (4): 485-504.
79. Yao S.J., Y.H. Song, L.Z. Zhang, and X.Y. Cheng. 2000. "Wavelet transform and neural networks for short-term electrical load forecasting." *Energy Conversion & Management* 41: 1975-1988.
80. Yu, L., S. Wang, and K.K. Lai. 2005. A novel nonlinear ensemble-forecasting model incorporating GLAR and ANN for foreign exchange rates." *Computers and Operations Research* 32: 2523-2541.
81. Zhang, G.P. 2003. "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing* 50: 159- 175.
82. Zhang, Q., and A. Benveniste. 1992. "Wavelet networks." *IEEE Transactions on Neural Networks*, 3(6): 889-898.