Predictive Accuracy of a Hybrid Generalized Long Memory Model for Short Term Electricity Price Forecasting

Souhir Ben Amor\textsuperscript{1,2} \quad Heni Boubaker\textsuperscript{3} \quad Lotfi Belkacem\textsuperscript{4}

Abstract

Accurate electricity price forecasting is the main management goal for market participants since it represents the fundamental basis to maximize the profits for market players. However, electricity is a non-storable commodity and the electricity prices are affected by some social and natural factors that make the price forecasting a challenging task. This study investigates the predictive performance of a new hybrid model based on Generalized long memory autoregressive model (k-factor GARMA), the Gegenbauer Generalized Autoregressive Conditional Heteroscedasticity (G-GARCH) process, Wavelet decomposition, and Local Linear Wavelet Neural Network (LLWNN) optimized using two different learning algorithms; the Back propagation algorithm (BP) and the Particle Swarm optimization algorithm (PSO). The performance of the proposed model is evaluated using data from Nord Pool Electricity markets. Moreover, it is compared with some other parametric and non-parametric models in order to prove its robustness. The empirical results prove that the proposed method performs well than other competing techniques.

Keywords: Wavelet decomposition, WLLWNN, k-factor GARMA, G-GARCH, Electricity spot price, Forecasting.

Research highlights

- The wavelet decomposition is used to improve the forecasting accuracy of the LLWNN model.
- A hybrid method is developed to forecast the electricity price.
- The hybrid method combines the parametric k-factor GARMA model, Wavelet decomposition, and the non-parametric LLWNN methods.
- The learning algorithms; BP and PSO methods are applied to obtain optimum WLLWNN structure to avoid overfitting.

---

\textsuperscript{1} Institute of high commercial studies of Sousse (IHEC), 40 Route de ceinture Sahloul III, 4054 Sousse, Tunisia
\textsuperscript{2} Corresponding author.
\textsuperscript{3}Institute of high commercial studies of Sousse (IHEC)
\textsuperscript{4}Institute of high commercial studies of Sousse (IHEC)
The Dual generalized k-factor GARMA-G-GARCH model is adopted for comparison purpose.

**Nomenclature**

- $\phi$, Father wavelet; $\psi$, Mother wavelet; $\varphi(x)$, is the active wavelet functions; $X(t)$, a linear combination at arbitrary level $J \in \mathbb{N}$ through different scales; $s_{j,k}$ and $d_{j,k}$ are the detail (fine scale) coefficients; $h_l$, $l = 0,\ldots,L - 1$ is the high-pass wavelet filter; $g_l$, $l = 0,\ldots,L - 1$ is the low pass scaling filter; $E$, The objective function to minimize of the Back propagation (BP) algorithm; $w_{i,j}$, represent the connection weight; $p$, is the number of input $(i = 1,2,\ldots,p)$ and $l$ is the number of the hidden units $(j = 1,2,\ldots,l)$; $r$, is the learning rate of the BP algorithm; $e$, Error between output values $\hat{y}$ by and real values $y$; $p_{\text{best}}$, the best value of each agent (in the PSO algorithm); $g_{\text{best}}$, the best Value of each agent in the group; $v_i$, is the velocity of agent $i$; $s_t$, is the current position of agent $i$; $a_i$ and $b_i$ are the scale and translation parameters, respectively, associated to the mother wavelet $\psi$; $\nu_i$, is the linear model associated to the LLWNN weights; $Y$, the output of the LLWNN model; $\lambda_m$, Gegenbauer frequency of the $k$-factor GARMA model; $\Phi(L)$ and $\Theta(L)$, Polynomials of the delay operator $L$, related to the $k$-factor GARMA model; $d_m$, Long-memory parameter associated with the $k$-factor GARMA model; $P_m(L)$, Gegenbauer polynomial of the $k$-factor GARMA model; $\mu_t$, the conditional mean of the time series; $e_t$, the residuals at time $t$ from the $k$-factor GARMA model; $f$, is a non-linear, non-parametric function determined by the neural network; $\sigma_t^2$, is the conditional variance; $\lambda$, Gegenbauer frequency of the $k$-factor GARMA model; $P_L(L)$, Gegenbauer polynomial of the G-GARCH model; $N$, is the number of observations of our sample.
1. Introduction

Since early 80s, there has been a radical change in the electricity market all across the world. The electricity market, which was largely functioning in a vertically integrated operating environment (where generation, transmission and distribution combined in one hand) shifted to a system where all these activities got separated and independent. The increasing privatization and deregulation of the market have also led to sharper competition (Amjady (2006)).

In power markets, price investigation has become an essential topic for all its participants. Background information about the electricity price is important for risk management. In fact, it may stand an advantage for a market player facing competition. Moreover, forecasting electricity prices at different time horizons is valuable for all industry stakeholders for cash flow analysis, capital budgeting, integrated resource planning, regulatory rule-making and financial procurement. Therefore, producers and consumers rely on price forecasting information to propose their corresponding bidding strategies.

However, the electricity prices has a behavior that is different from that of other commodities. The main difference is that electricity is a non-storable merchandize, accordingly a small variation in electricity production in hours or minutes can lead to huge price changes. Moreover, electricity prices show some particular characteristics such as high frequency, non-stationary behavior, multiple seasonality (on annual, weekly, and daily levels), calendar effect, high volatility, high percentage of unusual prices, hard nonlinear behavior, long memory, mean reversion, price spikes and limited information to the market participants, which may affect the prices dramatically (Weron (2006)). In this framework, due to the complexity of the electricity market, price forecasting has been the most challenging task. This has also motivated the researchers to develop efficient and intelligent methods to forecast the prices so that all stakeholders in the market can benefit out of it.

Herby, an appropriate forecast model for electricity prices should consider the previous mentioned features. Meanwhile, various methods from different modeling families are used for different analysis or planning issues. In the field of electricity price forecasting, broadly two methods; statistical or econometric time series models and soft computing models, these two approaches are found to have been applied.

In statistical models, Auto regressive integrated moving average ARIMA (Contreras et al. (2003)), and Generalized Autoregressive Conditional Heteroscedasticity GARCH (Garcia et al. (2005); Ghosh and Kanjilal (2014) and Girish. (2016)) are used extensively. Some authors combine the ARIMA and GARCH models to take into consideration the mean and variance of
the time series when modeling and forecasting the electricity price (Bowden and Payne (2008); and Heping and Jing (2013)). However, these models does not allow to take into consideration the long memory behavior that characterize the electricity prices. To overcome this limitation, some authors applied the Fractional Auto Regressive Moving Average (FARMA) model and the Fractionally Integrated Generalized Autoregressive Conditional Heteroscedasticity (FIGARCH) process, to model finite persistence in the conditional mean and the conditional variance, respectively (Koopman et al. (2007), Najeh et al. (2012)). In the spectral domain, these processes have a peak for very low frequencies close to the zero frequency. Hence, it is noteworthy 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. Woodward et al. (1998) generalizes the single frequency GARMA process to the so-called $k$-factor GARMA process that permits the spectral density function to be not just located at a single frequency but presented a finite number $k$ of frequencies in $[0, \pi]$, identified as the Gegenbauer frequencies or $G$-frequencies. The $k$-factor GARMA model has been applied by several authors to reproduce the seasonal patterns as well as the persistent effects in the stock markets (Boubaker and Sghaier (2015), Caporale and Gil-Alana (2014), Caporale et al. (2012)). Despite the compatibility of this model with the characteristics of electricity prices, few applications are oriented in the electricity market (Diongue et al. (2009), Soares and Souza (2006) and Diongue et al. (2004)).

However, the $k$-factor GARMA model allows only to model the conditional mean, assuming that the residuals are white noise with constant variance over time. However, this assumption is not proved in practice and the residuals are often characterized by a time-varying variance. To overcome this limitation, some researchers suggested to extend the $k$-factor GARMA model using the GARCH, FIGARCH and Fractionally Integrated Asymmetric Power ARCH (FIAPARCH) models (Boubaker (2015) Boubaker and Boutahar (2011) Boubaker and Sghaier (2015)). 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. The empirical evidence highlights the importance of modelling the periodic dynamics of the volatility. For this aim, Bordignon et al. (2007, 2010) suggested new category of GARCH models characterized by periodic long memory behavior. This category of
models introduces Gegenbauer polynomials into the equation of standard GARCH model, in the literature, only few studies apply this process for modelling the electricity spot price (Diongue et al. (2009)).

For the AI methods, application of artificial neural network (ANN) for the electricity price forecasting which considered as generalized periodic long-memory filters to estimate the time-varying volatility has become a research hotspots (Anbazhagan and Kumarappan (2014), Ioannis and Athanasios (2016), Harmanjot et al. (2016); and Jesus et al (2018)). Zhang and Benveniste (1992) suggested the Wavelet Neural Networks, which adopt a wavelet as an activation function, as an alternative to the classic sigmoidal functions. The WNs have been effectively used in the electricity prices forecasting in the short-term, (Bashir and El-Hawary (2000); Yao, Song, Zhang, and Cheng (2000); Gao and Tsoukalas (2001); Benaouda, et al. (2006); Ulugammai et al. (2007); Pindoriya et al (2009); and Mashud and Irena (2016)). Chen et al. (2004) developed a new type of wavelet neural network termed the Local Linear Wavelet Neural Network (LLWNN). 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)).

It is worth noting that a time series is often complicate in nature and an individual model is not able to detect the different patterns in the same way, in such case, adopting a hybrid methods or combining several models (Granger (1989)) has become a common practice to overcome the limitations of using a single model and enhance the forecasting accuracy.

In the literature, several combination methods have been suggested (Tseng et al. (2002); Zhang (2003); Taskaya and Casey (2005); Valenzuela et al. (2008); Khashei and Bijari (2010); Tan et al. (2010) and Jiang et al. (2017)).

Some researchers suggested that this combination can be improved using the wavelet decomposition. To exemplify, Shafie-khah et al. (2011) developed a novel hybrid method to forecast day ahead electricity price. This hybrid method is based on wavelet transform, Auto-Regressive Integrated Moving Average (ARIMA) models and Radial Basis Function Neural Networks (RBFN). Zhang et al (2017) proposed a hybrid approach that combines the wavelet transform, the kernel extreme learning machine (KELM) based on self-adapting particle swarm optimization and an auto regressive moving average (ARMA). Jinliang et al (2018) proposed a new hybrid model based on improved empirical mode decomposition (IEMD), autoregressive integrated moving average (ARIMA) and wavelet neural network (WNN) for Short term electricity load forecasting.
Our research focuses on resolving the issues of modeling and forecasting the feature of the electricity prices. In this framework, this paper provides three contributions. The main one is to improve the forecasting accuracy of LLWNN model. This objective is achieved through using the wavelet theory to decompose the historical price, instead of introducing it directly to the Network (in the input layer) and assess the effect of different levels of decomposition on forecasting accuracy. This technique allows the network to detect the existence of seasonal long memory behavior and thus better estimate the data, this novel network is termed WLLWNN. In fact, previous studies combined the wavelet decomposition with the ANN (Aggarwal et al. (2008)), this technique proves its performance but it’s not applied with the LLWNN. Meanwhile, some research (Pany (2011), Pany and Ghoshal (2013), Chakravarty et al. (2012)) assumed that this model can provide prediction accuracy since the characteristics associated to the electricity prices can be identified in the hidden layer by means of the wavelet activation function without needing any external decomposer/ composer. Secondly, a new hybrid model is proposed in order to exploit the strength of both parametric and non-parametric approaches, this model is called \( k \)-factor GARMA-WLLWNN process, which allows for long-memory behavior, associated with \( k \)-frequency and include a WLLWNN type model to describe time-varying volatility. Thirdly, in the literature of generalized long memory models, the authors adopt either the \( k \)-factor GARMA model or the G-GARCH model to estimate the conditional mean and the conditional variance of the time series, respectively. However, none of them considers the long memory and cyclical behavior simultaneously in the conditional mean and the conditional variance. In this study, with the aim of providing a robust tool for electricity prices prediction, we develop a new approach based on dual generalized long memory process using both the \( k \)-factor GARMA model and the G-GARCH model, which allows taking into account many stylized facts observed on the electricity spot prices, in particular stochastic volatility, long-range dependence and multi-seasonality behaviors. Moreover, through this study we aims at proving the performance of the proposed \( k \)-factor GARMA-WLLWNN model by comparing it with the \( k \)-factor GARMA-G-GARCH model. In fact, this step consist of testing the performance of the novel WLLWNN with a parametric model (G-GARCH) in modeling and forecasting the periodic long memory behavior that existing in the conditional variance.
The log-return of electricity price for the Nord Pool market is used in order to show the appropriateness and effectiveness of the proposed model to time series forecasting. The rest of the paper is organized as follows; in the next section, we present the econometric methodology, which includes the theoretical concepts of wavelet decomposition, the learning algorithms for optimizing the neural networks and the basic concept of the proposed hybrid $k$-factor GARMA-WLLWNN model and the $k$-factor GARMA-G-GARCH model. 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 LLWNN as well as the individual WLLWNN model, the hybrid $k$-factor GARMA-LLWNN model and the generalized long memory $k$-factor GARMA-G-GARCH model, and section 4 concludes the paper.

2. Econometric Methodology

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 and wavelets are good candidates for bringing out hidden patterns in price series (Nicolaisen et al. (2000)). In order 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.

Wavelets are used to analyze the time series through mathematical rules, where the time series is decomposed on a scale-by-scale basis. Hence, using a dilation and translation operations, this technique allow 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 denotes by $\phi$ and $\psi$, which are named father wavelet and mother wavelet, respectively, given by:

$$\int \phi(t) dt = 1,$$  \hspace{1cm} (1)

$$\int \psi(t) dt = 0.$$

(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:

\[ \phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k) \]  
\[ \psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \]

(3)  
(4)

Where \( j = 1, \ldots, J \) indexes the scale and \( k = 1, \ldots, 2^j \) indexes the translation. The parameter \( j \) is adopted as the dilation parameter of the wave’s functions. This parameter \( j \) adjusts the support of \( \psi_{j,k}(t) \) in order 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 \( J_0 \in \mathbb{N} \) through different scales of the type:

\[ X(t) = \sum_k s_{j_0,k} \phi_{j_0,k}(t) + \sum_{j \geq J} \sum_k d_{j,k} \psi_{j,k}(t) \]  

(5)

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 (fine scale) 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 (5) 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 MRA scheme. The recursive MRA scheme\(^5\), which is implemented by a two-channel filter bank (i.e. a high-pass wavelet filter \(\{h_l, \ l = 0, \ldots, L-1\}\)\(^6\) and its associated low pass scaling filter \(\{g_l, \ l = 0, \ldots, L-1\}\)\(^7\) satisfying the quadrature mirror relationship given by \(g_i = (-1)^{i+1} h_{L-1-l}\) for \(l = 0, \ldots, L-1\), 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) has constructed a class of wavelet functions where \(\phi\) is a function such that \(\{\phi(t-k), \ k \in \mathbb{Z}\}\) forms an orthonormal basis of piecewise constant functions of length one. The Daubechies wavelet has many desirable properties; its most useful property that is possessing the smallest support for a given number of vanishing moments.\(^8\) Daubechies defined a useful category of wavelet filters, termed the Daubechies compactly supported wavelet filters of width \(L\) and distinguishes between two choices; the extremal phase filters \(D(L)\) and the least asymmetric filters \(La(L)\).

### 2.2 Learning Algorithms for optimizing the neural networks

#### 2.2.1 The Back propagation Algorithm

For the BP algorithm, 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 network will learn quicker, but the learning process will

---

\(^5\) Mallat’s Multi-resolution Analysis is considered as a robust theoretical framework for critically sampled wavelet transformation (for more details, see Mallat (1989)).

\(^6\) The wavelet function (filter) of support \(L\) proceeds as a special filter possessing specific properties, such that (i) it integrates to zero, i.e., \(\sum_{l=0}^{L-1} h_l = 0\), (ii) has unit energy, i.e., \(\sum_{l=0}^{L-1} h_l^2 = 1\), and (iii) is orthogonal to its even shifts, i.e., \(\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{L-1} h_l h_{l+2n} = 0, \ \forall n \in \mathbb{N}^*\).

\(^7\) The scaling filter of support \(L\) is defined so as to satisfy the following properties, (i) \(\sum_{l=0}^{L-1} g_l = \sqrt{2}\), (ii) \(\sum_{l=0}^{L-1} g_l^2 = 1\), (iii) \(\sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{L-1} g_l g_{l+2n} = 0, \ \forall n \in \mathbb{N}^*\).

\(^8\) For Daubechies wavelets, the number of vanishing moments is half the filter length.
never converge if this rate is too high. In contrast, if the learning rate is so low, the network 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;

\[
E = \frac{1}{2} \sum (y_i - \omega_{i,0}\phi_i(x) - \omega_{i,1}p_i\phi_i(x) - \cdots - \omega_{i,q}\phi_i(x)\omega_{i,1,p_1}\phi_i(x) - \cdots - \omega_{i,q}\phi_i(x)\omega_{i,1,p_1}\phi_i(x) - \cdots - \omega_{i,q}\phi_i(x)\omega_{i,1,p_1}\phi_i(x))
\]  

(6)

Where \( y_i \) is the desired value, \( \phi(x) \) is the active wavelet functions, \( \omega_{i,0} \) represent the connection weight, \( p \) is the number of input \((i = 1,2,\ldots,p)\) and \( l \) is the number of the hidden units \((j = 1,2,\ldots,l)\). 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;

\[
\omega_{i+1} = \omega_i + \Delta \omega_i = \omega_i + \left( r \frac{\partial E}{\partial \omega_i} \right)
\]

(7)

Denote that \( r \) is the learning rate adopted in the WLLWNN model.

Where \( \frac{\partial E}{\partial \omega} \) for all weights are described by the following equations;

\[
\frac{\partial E}{\partial \omega_{i,j}} = \omega_{i,j} + r.e. \left( \frac{1}{2} \right) \left( x_i^2 + x_j^2 + \cdots + x_p^2 \right) \exp(-(x_1 - c_i)^2 + (x_2 - c_i)^2 + \cdots + (x_p - c_i)^2))
\]

(8)

For \( \forall j \neq 0; \)

\[
\frac{\partial E}{\partial \omega_{i,j}} = \omega_{i,j} + r.e. \left( \frac{1}{2} \right) \left( x_i^2 + x_j^2 + \cdots + x_p^2 \right) \exp(-(x_1 - c_i)^2 + (x_2 - c_i)^2 + \cdots + (x_p - c_i)^2)) \cdot x_j
\]

(9)

That is

\[
\frac{\partial E}{\partial \omega_{i,0}} = \omega_{i,0} + r.e. \left( \frac{1}{2} \right) \left( x_i^2 + x_j^2 + \cdots + x_p^2 \right) \exp(-(x_1 - c_i)^2 + (x_2 - c_i)^2 + \cdots + (x_p - c_i)^2))
\]

(10)

\[
\frac{\partial E}{\partial \omega_{i,1}} = \omega_{i,1} + r.e. \left( \frac{1}{2} \right) \left( x_i^2 + x_j^2 + \cdots + x_p^2 \right) \exp(-(x_1 - c_i)^2 + (x_2 - c_i)^2 + \cdots + (x_p - c_i)^2)) \cdot x_2
\]

(11)

\[
\frac{\partial E}{\partial \omega_{i,2}} = \omega_{i,2} + r.e. \left( \frac{1}{2} \right) \left( x_i^2 + x_j^2 + \cdots + x_p^2 \right) \exp(-(x_1 - c_i)^2 + (x_2 - c_i)^2 + \cdots + (x_p - c_i)^2)) \cdot x_3
\]

(12)

\[
\frac{\partial E}{\partial \omega_{i,3}} = \omega_{i,3} + r.e. \left( \frac{1}{2} \right) \left( x_i^2 + x_j^2 + \cdots + x_p^2 \right) \exp(-(x_1 - c_i)^2 + (x_2 - c_i)^2 + \cdots + (x_p - c_i)^2)) \cdot x_4
\]

(13)

Where \( e \) is the error between output values \( \hat{y} \) by and real values \( y \) \((e = \hat{y} - y)\); 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 proved clearly 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 is represented by \( vx \) and \( vy \). 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.

(a) The distance between current position and \( pbest \).

(b) The distance between the current position and \( gbest \).

The following equation is adopted to update the Velocity associated to each agent:

\[
v_i^{p+1} = w v_i^p + c_1 \text{rand}_i \ast (pbest_i - s_i^p) + c_2 \text{rand}_i (gbest - s_i^p) \tag{14}
\]

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:

\[
s_i^{p+1} = s_i^p + v_i^{p+1} \tag{15}
\]

The first term of equation (14), 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.

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 \( p_{best} \) of the agent, the \( p_{best} \) value is replaced by the current value. If the best value of \( p_{best} \) is better than the current \( g_{best} \), \( g_{best} \) 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 (14) and (15), 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 characterize the dimension of the search space. The optimization is finished when the personal best solution of each particle and the global best amount of the entire swarm are attended.

2.3 The Wavelet local linear wavelet neural WLLWNN

In this section, we propose a novel neural network based wavelet decomposition; this model consists into two steps. At the first step, the historical price data has been decomposed using Wavelet decomposition and introduced, in the second step, into the network (LLWNN) to produce the set of input variables, the formed model termed WLLWNN model.

The expression (5) represents the decomposition of \( X(t) \) into orthogonal components at different resolutions and constitutes the so-called wavelet multiresolution analysis (MRA). In the second step, this decomposed time series is shaped through the local linear wavelet neural network. The structure of WLLWNN model is shown in Figure.1.

For the LLWNN model, wavelets are used as an activation function generated from one single function \( \psi(x) \) by the operation of dilation and translation \( \psi(x) \).

\[
\psi(x) = \left\{ \begin{array}{l}
\psi_i = |a_i|^{-1/2} \psi \left( \frac{x - b_i}{a_i} \right) ; \quad a_i, b_i \in \mathbb{R}^n, i \in \mathbb{Z} \\
\end{array} \right.
\]

\[
x = (x_1, x_2, \ldots, x_n)
\]

\[
a_i = (a_{i1}, a_{i2}, \ldots, a_{in})
\]
\[ b_i = (b_{i1}, b_{i2}, \ldots, b_{in}) \]

\( \psi(x) \) is localized in both time space and the space scale, 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 \( u_i = w_{i0} + w_{i1}x_1 + \ldots + w_{in}x_n \) is introduced.

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

\[
Y = \sum_{i=1}^{m} (w_{i0} + w_{i1}x_1 + \ldots + w_{in}x_n)\psi_i(x)
\]  
(17)

The mother wavelet is

\[
\psi(x) = \frac{-x^2}{2} e^{\frac{-x^2}{2}}
\]  
(18)

\[
\psi(x) = e^{\left(\frac{x-a}{\sigma}\right)^2}
\]  
(19)

Where \( x = \sqrt{d_1^2 + d_2^2 + \ldots + d_n^2} \)  
(20)

---

Figure 1. The proposed Wavelet Local Linear Wavelet Neural Network
2.4 The k-factor GARMA model

Gray et al. (1989) proposed the \(k\)-frequency GARMA model, which is a general form of the ARFIMA model that allow the modeling of the periodic or quasi-periodic movement in the data. This model. The generalized autoregressive k-factor GARMA model is defined as follows;

\[
\Phi(L)\prod_{i=1}^{k} (I - 2\nu_{m,i}L + L^2)^{d_{m,i}} (y_t - \mu) = \Theta(L)\varepsilon_t, \tag{21}
\]

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 \(\nu_{m,i}\) provide information about periodic movement in the conditional mean, \(\varepsilon_t\) is a white noise perturbation sequence with variance \(\sigma^2\), \(k\) is a finite integer, \(|\nu_{m,i}| < 1, i = 1, 2, \ldots k\), \(d_{m,i}\) are long memory parameters of the conditional mean indicating how slowly the autocorrelations are damped, \(\mu\) is the mean of the process, \(\lambda_{m,j} = \cos^{-1}(\nu_{m,j}), i = 1, 2, \ldots k\), denote the Gegenbauer frequencies (G-frequencies).

The GARMA model with \(k\)-frequency is stationary when \(|\nu_{m,i}| < 1, d_{m,i} < 1/2\) or when \(|\nu_{m,i}| = 1\) and \(d_{m,i} < 1/4\), the model exhibits a long memory when \(d_{m,i} > 0\).

The main characteristic of model is given by the presence of the Gegenbauer polynomial;

\[
P_m(L) = \prod_{i=0}^{k} (I - 2\nu_{m,i}L - L^2)^{d_{m,i}} \tag{22}
\]

This polynomial may be considered a generalized long-memory filter that models the long-memory periodic behavior at \(k + 1\) different frequencies. When thinking of the \(\lambda_i\) as the driving frequencies of a cyclical pattern of length \(S\), \(\lambda_{m,j} = (2\pi j / S)\), and \(k+1 = [S/2] + 1\), where \([\cdot]\) stands for the integer part. To highlight the contribution of \(P_m(L)\) at frequencies \(\lambda = 0\) and \(\lambda = \pi\), equation (22) can be written as:

\[
P_m(L) = (I - L)^{d_{m,k}} (I + L)^{d_{m,L}} \prod_{i=1}^{k+1} (I - 2\nu_{m,i}L - L^2)^{d_{m,i}} \tag{23}
\]

Where \(I(E) = 1\) if \(S\) is even and zero otherwise and \(k+1 = [S/2] + 1 - I(E)\).

For a GARMA model with a single frequency, when \(\nu = 1\), the model is reduced to an ARFIMA \((p,d,q)\) model, and when \(\nu = 1\) 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 $\lambda = \cos^{-1}(\nu)$, which varies in the interval $[0, \pi]$. It is important to note that when $|\nu| < 1$.

### 1.1 The Gegenbauer-GARCH model

The generalized long memory GARCH or Gegenbauer-GARCH (G-GARCH) introduced by Bordignon et al. (2007, 2010), generalized the FIGARCH and the Fractionally Integrated Exponential GARCH (FIEGARCH) models by allowing the periodic long memory patterns of the power spectrum in both the zero and non-zero frequencies and then its able to estimate the seasonality in different frequencies in the conditional variance.

The G-GARCH model for conditional variance modeling is given by;

$$\sigma_t^2 = \gamma + \beta(L)\sigma_t^2 + \{I - \beta(L) - \left[(I-L)^{d_{\nu,0}}(I+L)^{d_{\nu,1}} \prod_{i=1}^{k=1} (I-2\nu_{\nu,i}L + L^2)^{d_{\nu,i}}\right] \psi(L)\} \epsilon_t^2$$  \hspace{1cm} (24)

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

When $d_{\nu,0} = d_{\nu,1} = \ldots = d_{\nu,k}$, all the involved frequencies have the same degree of memory.

Bordignon et al. (2007) proposed to model the logarithm of the conditional variances in order to guarantee the conditional variance positivity. For this purpose, a filter to a generalized log-GARCH model is proposed;

$$P_{\nu}(L)\psi(L)[\ln(\sigma_t^2) - \tau] = \gamma + [I - \beta(L)]\vartheta_t$$  \hspace{1cm} (25)

Where $P_{\nu}(L)$ is the generalized long memory filter introduced into a GARCH structure, $\vartheta_t = \ln(\epsilon_t^2) - \tau - \ln(\sigma_t^2)$ is a martingale difference and $\tau = \mathbb{E}[\ln(\epsilon_t^2)]$. The expected value depends on the distribution of the idiosyncratic shock and ensures that $\vartheta_t$ is a martingale difference, given that $\ln(\epsilon_t^2) = \ln(\sigma_t^2) + \ln(\epsilon_t^2)$. Under the Gaussian assumption $\tau = -1.27$.

The expression for conditional variance implied by (24) is:

$$\ln(\sigma_t^2) = \gamma + \beta(L)\ln(\sigma_t^2) + [I - \beta(L) - P_{\nu}(L)\psi(L)](\ln(\epsilon_t^2) - \tau)$$  \hspace{1cm} (26)
Since we are modelling $\ln(\sigma^2_t)$ instead of $\sigma^2_t$, no constraints for variance positivity are necessary.

The $k$-factor GARMA model extended using the Gegenbauer GARCH model form the proposal dual generalized $k$-factor GARMA-G-GARCH model that is able to capture seasonality and long memory dependence in both the conditional mean and the conditional variance (see Figure 2).

Concerning the estimation of the $k$-factor GARMA-G-GARCH model, we adopt an estimation procedure based on the wavelets following the methodology proposed by Whitcher (2004).

In this paper we focus on the discrete wavelet packet transform (DWPT), which deals with the existence of seasonalities and allows to decorrelate the spectrum of the process (see Boubaker (2015) and Guegan and Lu (2009), for more details).

1.2 The proposed $k$-factor GARMA-WLLWNN hybrid model

Our hybrid methodology combines a $k$-factor GARMA model and the proposed WLLWNN model. The $k$-factor GARMA model offers greater flexibility in modeling simultaneous short and long-term behavior of a seasonal time series. In addition, the choice of WLLWNN in our hybrid model is motivated by the wavelet decomposition and its local linear modeling ability. It may be reasonable to consider time series to be composed by two components; the first one is a parametric form with unknown parameters where a parametric method seems appropriate for such processes. The second component related to the residuals, this part usually present no specific process. Hence, it is difficult to determine the appropriate model that can be deal with this part of the time series. For this reason, a non-parametric model seems appropriate for modelling the residuals. This choice is motivated by the fact that the non-parametric model can reduce modelling bias by imposing no specific model structure, rather than certain smoothness assumption, and therefore, they are particularly used when we have little information or when we want to be flexible about the underling model.

Our methodology consists into two steps; in the first step, the aim is to model the conditional mean using a $k$-factor GARMA model. However, 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 (see Figure.1).

Hence, a time series can be written as:
\[ y_t = \mu_t + \varepsilon_t \]  
\[ t \]

Where \( \mu_t \) denote the conditional mean of the time series, and \( \varepsilon_t \) is the residuals. The \( k \)-factor GARMA model is used to reproduce the conditional mean \( \mu_t \) using equation (21).

In the second step, the residuals from the parametric model \( (\varepsilon_t) \) are used as a proxy for the corresponding volatility and modeled using the WLLWNN model.

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

\[ \varepsilon_t = y_t - \hat{\mu}_t \]  
\[ t \]

Where \( \hat{\mu}_t \) is the forecast value from the estimated relationship (equation 21). Thereafter, 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:

\[ \varepsilon_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-n}) \]  
\[ t \]

Where each \( \varepsilon_{t-i} \) is decomposed using the Wavelet Transform (equation 5), \( f \) 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 17) gives the forecasting results:

\[ \hat{y}_t = \hat{\mu}_t + \hat{\varepsilon}_t \]  
\[ t \]

Hence, this global prediction, represent the result of forecasting both, the conditional mean and the conditional variance of the time series (see Figure 2).

To conclude, the proposed hybrid model exploits the originality and the strength of the \( k \)-factor GARMA model as well as the WLLWNN model to detect the different features existing in the data in order to benefit the complementary characteristics of the models from which there are composed. Thus, the proposed hybrid model can be an efficient way giving a more general and accurate method than other hybrid models.
2. Empirical Methodology

2.1 Data description and Preliminary study

The predictive performance of the proposed model is tested using hourly spot prices from the Nord Pool electricity market, our sample covers the period between 1st of January 2016 and 30 June 2017, in total $T = 13128$ hourly observations, illustrated in Figure 3. In this study, we consider the logarithm of these series ($R_t = \Delta LogP_t$) to make the series stationary.
As shown in Figure 4, for the RSP, the spectral density, traced by the periodogram, shows several peaks at equidistant frequencies, which proves the presence of several seasonalties. Using the GPH (Geweke and Porter-Hudak, (1983)) and LW (Robinson, (1995)) statistics, we tested the long-range dependence in the conditional mean. Corresponding results shown in Table 1 indicate evidence of long memory.

### Table 1. Results of GPH and LW long-range dependence tests in the conditional mean

| Bandwidth | RSP N=13128 | GPH | LW |
|------------|-------------|-----|-----|
|            |             | $\hat{d}_m$ | Standard error | p-value | $\hat{d}_m$ | Standard error | p-value |
| $N^{0.5}=115$ | -0.3432*** | 0.0642 | 0.0000 | -0.4554*** | 0.0466 | 0.0000 |
| $N^{0.6}=296$ | -0.2858*** | 0.0386 | 0.0000 | -0.3473*** | 0.0290 | 0.0000 |
| $N^{0.7}=764$ | -0.4354*** | 0.0236 | 0.0000 | -0.6390*** | 0.0180 | 0.0000 |
| $N^{0.8}=1970$ | -0.3698*** | 0.0146 | 0.0000 | -0.5210*** | 0.0112 | 0.0000 |

*Note: Asterisks denote significance at ***1% level.

### 2.2 Estimation Results

- Conditional mean estimation results (the $k$-factor GARMA model)

### Table 2. Estimation of the k-factor GARMA model: a wavelet based approach

| Parameters | $k$-factor GARMA model estimation |
|------------|----------------------------------|
| $\hat{\Phi}$ | 0.0213*** |
| $\hat{\Theta}$ | - |
| $\mu$ | - |
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 4), are unbounded at equidistant frequencies, which proves presence of several seasonalities. They show special peaks at frequencies $\hat{\lambda}_{m,1} = 0.0416$ ($T = 24.038 \approx 24$ hours = 1 day), $\hat{\lambda}_{m,2} = 0.0834$ ($T = 11.99 \approx 12$ hours $\approx 1/2$ day), and $\hat{\lambda}_{m,3} = 0.1713$ ($T = 5.83 \approx 6$ hours $= 1/3$ day), corresponding to cycles with daily, semi-daily and third-daily periods, respectively.

The second step consist in modeling 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, in order to select the adequate method.

- **Conditional variance estimation results**
  - **The WLLWNN estimation results**

In the NNs, the data training allows the model to develop a generalized structure, which is able to estimate accurately the unknown values. In order to create the training and testing data sets, it is essential to select the input variables. In reality, a number of factors like historical prices, bidding strategies, weather, fuel prices and demand elasticity, affect electricity prices. However, the process will be complicated, if we include all these factors in the forecasting model, since we need a separate module of relevant feature selection, in order to identify and include the most suitable input features. Furthermore, the redundancy in the input features can reduce the performance of the model instead of improving it (Kohavi and George (1997); and Pindoriya et al. (2008)). In order to conserve the simplicity of our methodology and for comparison
purposes, we have used historical prices as the unique input variable referring to several researches. Hence, residuals resulting from the $k$-factor GARMA modeling are considered here as the input of the novel WLLWNN in order 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;

$$y_{\text{norm}} = \frac{y_{\text{org}} - y_{\text{min}}}{y_{\text{max}} - y_{\text{min}}}$$

(31)

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

**Wavelet decomposition**

These normalized data are then decomposed using the MODWT\(^9\) with Daubechies least asymmetric ($L\alpha$) wavelet filter of length $L = 8(L\alpha(8))^{10}$. This wavelet filter has been frequently adopted in the financial literature and it has been proved that $L\alpha(8)$ provides the best performance for the wavelet time series decomposition. Our MODWT decomposition goes up to level $J = 13$ 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 Gencay et al. (2002)) plotted in Figures 5 and 6. The time series is decomposed into 13 details ($D_j(1), ... D_j(13)$). It has been proved empirically that the level of decomposition 10 of price series (Figure 5) is optimum for an accurate forecasting. Figure 6 represents the wavelet smooth S14, it is noticed that the deterministic trend is nonlinear.

---

\(^9\)The wavelet transform adapts itself intelligently to capture patterns through a wide range of frequencies and hence, has the capacity to detect events that are local in time. Therefore, the wavelet transform are considered as a perfect tool to analyze non-stationary time series. For the purpose of measuring the local regularity of a signal, it is not necessary to adopt a wavelet with a narrow frequency support, but vanishing moments are essential. If the wavelet has M vanishing moments, then we prove that the wavelet transform can be interpreted as a multiscale differential operator of order M, i.e. The first M moments of the wavelet coefficients are zero; $\int t^{(r)} \varphi(t) dt = 0, \quad r = 1, ..., M$. This gives a first relation between the differentiability of the time series and its wavelet transform decay at fine scales. According to Gencay et al. (2002) the Daubechies wavelet filter guarantees the resulting DWT coefficients be stationary and hence, protect ourselves from problems caused by a non-stationary time series.

\(^{10}\) 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. Since the width of the wavelet filter increases, it can better match to the features of the time series. 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 Gencay et al. (2002).
The dataset is divided into three successive parts as follows: (a) A sample of 2000 observations to initialize the network training, (b) a training set and (c) a test set. The forecasting experiment is performed over the test set using an iterative forecasting scheme, the models are tested for five time horizons; 6, 12, 24, 48 and 72 hours ahead. Details of the datasets are given in the Table 3.

Table 3. Details of datasets

| Total sample | Initialization Sample | Training Sample | Testing Sample |
|--------------|-----------------------|-----------------|----------------|
| $N = 13128$  | From 01/01/2016 To 24/03/2016; 2000 observations | From 25/03/2016 To 27/06/2017; 11056 observations. | From 28/06/2017 to 30/06/2017; 72 observations (divided into 5 time horizons) |

In order to find the best neural network architecture the Back Propagation algorithm (BP) and the Particle Swarm Optimization Algorithm (PSO) are applied to optimize the parameters. Table 4 and Table 5 provide the summary of information related to the network architecture.

Table 4. LLWNN based BP Algorithm architecture

| Number of hidden layers | 10 |
| Learning rate           | 0.5 |
| Layer conversion function | Wavelet Function |
| Algorithm               | Back Propagation (BP) Learning Algorithm |
Parameters adopted for running PSO are presented as follow;

### Table 5. LLWNN based PSO Algorithm architecture

| Parameter                     | Value   |
|-------------------------------|---------|
| Number of population          | 20      |
| Number of generation          | 200     |
| $C_1, C_2$                    | 1.05    |
| Maximum velocity              | 1       |
| Minimum velocity              | 0.3     |
| Number of hidden layers       | 10      |
| Learning rate                 | 0.5     |
| Layer Activation function     | Wavelet Function |

![Figure 7: Periodogram of the residuals of the $k$-factor GARMA model](image)

- **The $k$-factor GARMA-G-GARCH estimation results**

As shown in Figure 7, the presence of several peaks at equidistance frequencies in the spectral densities proves the presence of several seasonalities. Note that the residuals from the $k$-factor GARMA model are used as a proxy of the corresponding volatility, on which we performed Long memory tests. As reported in Table 6, the results of the GPH and LW indicate the presence of long memory in the conditional variance.
Table 6. Results of GPH and LW long-range dependence tests in the conditional variance

| Bandwidth | GPH | LW |
|-----------|-----|----|
|           | $\hat{d}_v$ | Standard error | p-value | $\hat{d}_v$ | Standard error | p-value |
| RSP       |     |     |     |     |     |     |
| $N^{0.5} = 115$ | 0.4276*** | 0.0642 | 0.0000 | 0.3440*** | 0.0466 | 0.0000 |
| $N^{0.6} = 296$ | 0.4171*** | 0.0386 | 0.0000 | 0.3491*** | 0.0290 | 0.0000 |
| $N^{0.7} = 764$ | 0.2979*** | 0.0236 | 0.0000 | 0.2623*** | 0.0180 | 0.0000 |
| N=13128   | 0.2035*** | 0.0146 | 0.0000 | 0.1757*** | 0.0113 | 0.0000 |

Note: Asterisks denote significance at ***1% level.

Table 7. The $k$ -factor GARMA-G-GARCH Estimation Results

| $k$-factor GARMA model estimation | The G-GARCH model estimation |
|----------------------------------|-------------------------------|
| $\hat{\phi}$                     | $\hat{\psi}$                 |
| $\hat{\Theta}$                   | $\hat{\beta}$                |
| $\mu$                            | $\hat{\gamma}$               |
| $\hat{d}_{m,1}$                  | $\hat{d}_{v,1}$              |
| $\hat{d}_{m,2}$                  | $\hat{d}_{v,2}$              |
| $\hat{d}_{m,3}$                  | $\hat{d}_{v,3}$              |
| $\hat{\lambda}_{m,1}$            | $\hat{\lambda}_{v,1}$        |
| $\hat{\lambda}_{m,2}$            | $\hat{\lambda}_{v,2}$        |
| $\hat{\lambda}_{m,3}$            | $\hat{\lambda}_{v,3}$        |

Note: Asterisks denote significance at ***1% level.

The residuals from the $k$ -factor GARMA are modeled using the G-GARCH model in order 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 7. The spectral density, represented by periodogram (Figure 7), is unbounded at equidistant frequencies, which proves the presence of several seasonalities. They show special peaks at frequencies $\lambda_{v,1} = 0.0187$ (T=53h, 28min $\approx$ 2 days), $\lambda_{v,2} = 0.0413$ (T=24h, 12min $\approx$ 1 day), and $\lambda_{v,3} = 0.0982$ (T=10h, 10min $\approx$ 1/2 day), that corresponding to cycles with two days, daily and semi-daily periods, respectively.
2.3 Forecasting Results comparaison

This section is devoted to the evaluation of the estimated models in a multi-step-ahead forecasting task. Since forecasting is basically an out-of-sample problem, we prefer to apply out-of-sample criteria. Accordingly, five different periods (6 hours, 12 hours, one day, two days, and three days) were selected in aim to ensure the robustness of forecasting results. In order to evaluate the forecasting accuracy, we apply three evaluation criteria, namely the Mean Absolute error (MAE), the Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE), given respectively by:

\[
MAE = \frac{1}{N-t_i} \sum_{t=i}^{N} \left| y_{t+h} - \hat{y}_{t+h} \right|
\]

\[
MSE = \frac{1}{N-t_i} \sum_{t=i}^{N} \left( y_{t+h} - \hat{y}_{t+h} \right)^2
\]

\[
RMSE = \left( \frac{1}{N-t_i} \sum_{t=i}^{N} \left( y_{t+h} - \hat{y}_{t+h} \right)^2 \right)^{1/2}
\]

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

Table 8. Out of sample Forecasts Results

| Models                        | Criterion | \( h = 6 \)   | \( h = 12 \)  | \( h = 24 \)  | \( h = 48 \)  | \( h = 72 \)  |
|-------------------------------|-----------|---------------|---------------|---------------|---------------|---------------|
| LLWNN based BP Algorithm      | MAE       | 0.0078        | 0.0104        | 0.0169        | 0.0158        | 0.0101        |
|                               | MSE       | 7.047×10^{-5} | 1.4995×10^{-4} | 5.0542×10^{-4} | 3.6611×10^{-4} | 1.6566×10^{-4} |
|                               | RMSE      | 0.0084        | 0.0122        | 0.0225        | 0.0191        | 0.0129        |
| LLWNN based PSO Algorithm     | MAE       | 0.0082        | 0.0116        | 0.0135        | 0.0137        | 0.0130        |
|                               | MSE       | 1.0331×10^{-4} | 2.4644×10^{-4} | 2.6787×10^{-4} | 3.1191×10^{-4} | 2.9709×10^{-4} |
|                               | RMSE      | 0.0102        | 0.0157        | 0.0164        | 0.0177        | 0.0172        |
| The hybrid k-factor           | MAE       | 0.0038        | 0.0081        | 0.0087        | 0.0101        | 0.0089        |
| GARMA-LLWNN based BP algorithm| MSE       | 3.1983×10^{-5} | 9.5044×10^{-5} | 1.2492×10^{-4} | 1.6142×10^{-4} | 1.2499×10^{-4} |
|                               | RMSE      | 0.0057        | 0.0097        | 0.0112        | 0.0127        | 0.0112        |
The hybrid k-factor GARMA-LLWNN based PSO algorithm

| MAE           | 0.0016          | 0.0023          | 0.0057          | 0.0061          | 0.0076          |
|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| MSE           | 3.5028 × 10⁻⁶  | 3.6221 × 10⁻⁶  | 2.2690 × 10⁻⁵  | 2.5612 × 10⁻⁵  | 3.9241 × 10⁻⁵  |
| RMSE          | 0.0018          | 0.0019          | 0.0047          | 0.0050          | 0.0062          |

WLLWNN based BP Algorithm

| MAE           | 4.6968 × 10⁻⁶  | 5.1739 × 10⁻⁶  | 4.0551 × 10⁻⁶  | 1.8927 × 10⁻⁶  | 4.5983 × 10⁻⁶  |
|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| MSE           | 2.2306 × 10⁻¹¹ | 2.7097 × 10⁻¹¹ | 1.7424 × 10⁻¹¹ | 4.4187 × 10⁻¹² | 2.6346 × 10⁻¹¹ |
| RMSE          | 4.7229 × 10⁻⁶  | 5.2055 × 10⁻⁶  | 4.1741 × 10⁻⁶  | 2.1021 × 10⁻⁶  | 5.1329 × 10⁻⁶  |

WLLWNN based PSO Algorithm

| MAE           | 3.7001 × 10⁻⁸  | 7.2902 × 10⁻⁸  | 1.6286 × 10⁻⁷  | 5.9591 × 10⁻⁸  | 2.7986 × 10⁻⁷  |
|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| MSE           | 1.7781 × 10⁻¹⁵ | 6.7075 × 10⁻¹⁵ | 3.3395 × 10⁻¹⁴ | 7.7661 × 10⁻¹⁵ | 9.8627 × 10⁻¹⁴ |
| RMSE          | 4.2170 × 10⁻⁸  | 8.1899 × 10⁻⁸  | 1.8274 × 10⁻⁷  | 8.8121 × 10⁻⁸  | 3.1405 × 10⁻⁷  |

The hybrid k-factor GARMA-WLLWNN based BP Algorithm

| MAE           | 2.0409 × 10⁻⁶  | 1.7647 × 10⁻⁶  | 2.7731 × 10⁻⁶  | 3.6631 × 10⁻⁶  | 6.4121 × 10⁻⁷  |
|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| MSE           | 4.3281 × 10⁻¹² | 3.3223 × 10⁻¹² | 8.0457 × 10⁻¹² | 1.3885 × 10⁻¹¹ | 4.2383 × 10⁻¹³ |
| RMSE          | 2.0804 × 10⁻⁶  | 1.8227 × 10⁻⁶  | 2.8365 × 10⁻⁶  | 3.7263 × 10⁻⁶  | 6.5102 × 10⁻⁷  |

The hybrid k-factor GARMA-WLLWNN based PSO Algorithm

| MAE           | 4.7726 × 10⁻⁹  | 2.9456 × 10⁻⁸  | 3.1504 × 10⁻⁸  | 2.7194 × 10⁻⁸  | 2.9496 × 10⁻⁸  |
|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| MSE           | 3.6761 × 10⁻¹⁷ | 1.0776 × 10⁻¹⁵ | 1.8611 × 10⁻¹⁵ | 1.2513 × 10⁻¹⁵ | 1.4299 × 10⁻¹⁵ |
| RMSE          | 6.0631 × 10⁻⁹  | 3.2826 × 10⁻⁸  | 4.3140 × 10⁻⁸  | 3.5374 × 10⁻⁸  | 3.7813 × 10⁻⁸  |

The k-factor GARMA-GARCH model

| MAE           | 5.2316 × 10⁻⁷  | 6.7428 × 10⁻⁷  | 4.8637 × 10⁻⁶  | 3.6238 × 10⁻⁸  | 3.2478 × 10⁻⁸  |
|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| MSE           | 3.2458 × 10⁻¹⁶ | 8.5214 × 10⁻¹⁴ | 2.9452 × 10⁻¹⁴ | 4.5831 × 10⁻¹⁵ | 2.9773 × 10⁻¹⁵ |
| RMSE          | 1.8000 × 10⁻⁸  | 2.9191 × 10⁻⁷  | 1.7160 × 10⁻⁷  | 6.7700 × 10⁻⁸  | 5.4600 × 10⁻⁸  |

In order to evaluate the prediction performance of the proposed hybrid methodology, this paper has taken into account five models: the individual LLWNN model, the hybrid k-factor GARMA-LLWNN model, the proposed WLLWNN model, the hybrid k-factor-GARMA-WLLWNN model and the k-factor GARMA-GARCH model, we applied two different learning algorithms (BP and PSO) for the training of the networks. Moreover, we adopted five time horizons; 6 hours, 12 hours, one day, 2 days and 3 days ahead forecasting, using the MAE, the MSE and the RMSE out of sample criteria. The forecast evaluation results are reported in Table 8. They show that:

- The individual LLWNN based PSO algorithm outperforms the individual LLWNN based BP algorithm, in addition, the individual WLLWNN based PSO algorithm outperforms the individual WLLWNN based BP algorithm; these results 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 presents some drawbacks such as very slow convergence to a local minimum. However, in the case of training with PSO algorithm, weights are characterized by particle position. These particles velocity and position are updated, in order to search for personal best and global best values. This will avoid the convergence of weights to a local minimum.

- The hybrid $k$-factor GARMA-LLWNN model outperforms the individual LLWNN model, hence, the individual LLWNN model is unable to detect, to model and to predict the features existing in the electricity prices, meanwhile, when it is compared with the hybrid model, this last one provides prediction that is more accurate. Consequently, this network need an external filter to better estimate the data since the adoption of the $k$-factor GARMA model in the first step enhance the results of forecasting. This result contradicts existing literature; Pany (2011), Pany and Ghoshal (2013) and Chakravarty et al. (2012) proving that the use of individual LLWNN models for forecasting electricity prices provides the best prediction results compared to other ordinary networks such as multi-layer perceptron networks (MLP). This is because the features of the electricity prices can be detected by the wavelet basis activation function for hidden layer neurons without any external decomposer/composer and not having too many hidden units. In addition, this result confirms the effectiveness of the hybrid model in improving the forecasting accuracy.

- The $k$-factor GARMA-G-GARCH model outperforms the hybrid $k$-factor GARMA-LLWNN model in terms of prediction accuracy. This can be explained by the fact that the $k$-factor GARMA-G-GARCH model takes into account the seasonal long-memory in both the conditional mean and the conditional variance, making this model a robust tool that can deal with the features of the electricity prices. Furthermore, despite its capacity as a nonlinear, nonparametric model, and its particularity by having a wavelet activation function and local linearity, the LLWNN model is unable to predict the seasonality and the long memory dependence in the data. Since, when it is compared with the G-GARCH model, this last one provides a prediction results that is more accurate. This is explained by the ability of the G-GARCH in modelling the seasonal long memory behaviour in the conditional variance. This also proves the importance of taking into account the seasonal long memory behaviour to enhance forecasting accuracy (Nowotarski and Weron (2016)).

- The novel WLLWNN overcomes the limitation of the LLWNN that is related to the inability of the network to detect and model the periodic long memory behaviour in the data,
since it shows its effectiveness when we compare it with the \( k \)-factor GARMA-G-GARCH and the \( k \)-factor GARMA-LLWNN model.

- As shown in Figures 8 the predictions of the \( k \)-factor GARMA-WLLWNN model based PSO algorithm for all the five horizons are very close to the real values (see Figure.19). That is confirm the forecasting results (Table 10), which indicate that the \( k \)-factor GARMA-WLLWNN model prediction errors are the smallest for all evaluation criteria and for all forecast time horizons.

To sum up, 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 several seasonalities. 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 dealing with the features of the electricity prices and provide the best forecasting results.
Electricity price forecasting is crucial in the decision making of power market participants. Developing more accurate price forecasting methods has become an important research topic. In this paper, we develop a relatively novel neural network model, termed Wavelet Local Linear Wavelet Neural Network (WLLWNN). In this model, the historical price data has been decomposed into the wavelet domain constitutive sub-series using Wavelet Transform and then introduced into the network to produce the set of input variables and form the WLLWNN model. This novel network is claimed to have a higher generalization performance than the LLWNN. 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 BP algorithms update weights in the direction of the negative gradient. NNs training with BP algorithms presents certain drawbacks such as very
slow convergence that can be trapped in the local minimum. However, weights in PSO algorithm are represented by particle position. This particle velocity and position are updated, with the aim to search for personal best and global best values. This will avoid the convergence to a local minimum. Our experimental results showed the superiority of PSO algorithm. Moreover, the effectiveness of the novel WLLWNN is further enhanced by coupling it with the semi-parametric $k$-factor GARMA model to develop a hybrid model termed as $k$-factor GARMA-WLLWNN model. In this way, the unique strength of each method is combined to detect different patterns in the data. The performance of the proposed model is evaluated using data from Nordic Electricity market and compared with the dual generalized long memory $k$-factor GARMA-G-GARCH model, the hybrid $k$-factor GARMA-LLWNN model, and the individual WLLWNN. 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.

To recapitulate, in addition of developing an efficient method for price forecasting, we presented a detailed analysis of several price-forecasting techniques. Our proposed hybrid model shows its performance; moreover, it is suitable and feasible to real world application since it exploits information that is available to the public (historical prices) that can be easily collected by all market participant before submitting their bids in the market and gives to investors enough time to respond.

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] Amjady, N. 2006. “Day Ahead Price Forecasting of Electricity Markets by a New Fuzzy Neural Network”. IEEE, Transaction on Power SystA3: A108ems.4: No.2.

[4] Anbazhagan, S., Kumararappan. 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.

[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] Bowden, N., Payne, J.E., 2008. Short term forecasting of electricity prices for MISO hubs: evidence from ARIMA-EGARCH models. Energy Economics. 30, 3186-3197.
[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] Bordignon, S., M. Caporin, and F. Lisi. 2007. “Generalised long-memory GARCH models for intra-daily volatility”. Computational Statistics & Data Analysis 51: 5900-5912.
[11] Bordignon, S., M. Caporin, and F. Lisi. 2010. “Periodic long memory GARCH models” Econom Rev 28:60-82.
[12] 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.
[13] Boubaker. H. 2015. “Wavelet Estimation of Gegenbauer Processes: Simulation and Empirical Application”. Comput Econ 46 :551-574.
[14] Boubaker; H., and M. Boutahar. 2011. “A wavelet-based approach for modelling exchange rates”. Stat Methods Appl 20: 201-220.
[15] 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.
[16] Caporale, G.M., L.A. Gil-Alana. 2014. “Long-run and cyclical dynamics in the US stock market”. J. Forecast 33: 147-161.
[17] Caporale, GM., 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.
[18] Chakravarty, S., M. Nayak, and M. Bisi. 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.
[19] Chen, Y., J. Dong, B. Yang, and Y. Zhang, 2004. “A local linear wavelet neural network.” Intelligent Control and Automation, IEEE: 1954-1957.
[20] Cheung, Y.W. 1993. “Long memory in foreign-exchanges rates.” J Bus Econ Stat 11: 93-101.
[21] 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.
[22] Daubechies, I. 1992. “Ten lectures on wavelets” Philadelphia: SIAM.
[23] Diongue, A.K., G. Dominique, and V. Bertrand. 2004. “A k-factor GIGARCH process: estimation and application on electricity market spot prices.” Probabilistic Methods Applied to Power Systems: 12-16.
[24] 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.
[25] Gao, R., and H. I. Tsoukalas. 2001. “Neural-wavelet methodology for load forecasting.” Journal of Intelligent & Robotic Systems 31: 149-157.
[26] 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.
[27] Gencay, R., F. Selcuk, and B. Whitcher. 2002. “An Introduction to Wavelets and Other Filtering Methods in Finance and Economics”. Academic-Press.
[28] 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.

[29] 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.

[30] Granger, C.W.J. 1989. “Combining forecasts: twenty years later”. *International Journal of Forecasting*; 8:167-73.

[31] Gray, H.L., N.F. Zhang, and W.A. Woodward. 1989. “On generalized fractional processes”. *J. Time Ser. Anal* 10 (3): 233-257.

[32] Gue Guegan, D., Lu, Z., 2009. Wavelet method for locally stationary seasonal long memory processes.

[33] Heping, L and Jing., S. 2013. “Applying ARMA–GARCH approaches to forecasting short term electricity prices”. *Energy Economics, Volume 37, Pages 152-166.*

[34] 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.*

[35] Jinliang Z., Yi-Ming W., Dezhi. L., Zhongfu. T., Jianhua. Z. 2018. Short term electricity load forecasting using a hybrid model. *Energy, V 158, P 774-781*

[36] Kennedy, J.R. and Eberhart. 1995. “Particle swarm optimization”. *IEEE international conference on neural networks. 4: 1942-1948.*

[37] Khashei, M. and M. Bijari. 2010. “An artificial neural network model for time series forecasting.” *Expert Syst. Appl. 37: 479-489.*

[38] Kohavi, R., H. J. George. 1997. “Wrappers for feature subset selection”. *Artificial Intelligence 97 (1-2): 273-324.*

[39] Mallat, S. 1989. “A theory for multiresolution signal decomposition: The wavelet representation”. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11: 674-693.*

[40] 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.*

[41] Nowotarski, J., and R. Weron. 2016. “On the importance of the long-term seasonal component in day-ahead electricity price forecasting.” *Energy Economics 57: 228-235.*

[42] Pany, P.K. 2011. “Short-Term Load Forecasting using PSO Based Local Linear Wavelet Neural Network.” *International Journal of Instrumentation 1: Issue-2.*

[43] 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.*

[44] Pindoriya, N.M., S.N. Singh, and S.K. Singh. 2008. “An adaptive wavelet neural network-based energy price forecasting in electricity markets”. *IEEE Trans Power Syst 23 (3): 1423-1432.*

[45] Robinson, P.M. 1995. “Log-Periodogram Regression of Time Series with Long-Range Dependance.” *Annals of statistics 23: 1048-1072.*

[46] Soares, L.J., and L.R. Souza. 2006. “Forecasting electricity demand using generalized long memory.” *International Journal of Forecasting 22, 17-28.*
[47] 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.

[48] Taskaya. T., and M.C. Casey. 2005. “A comparative study of autoregressive neural network hybrids.” *Neural Networks* 18: 781-789.

[49] 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.

[50] Ulugammi, 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.

[51] 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.

[52] Weron, R. 2006. “Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach”. *The Wiley Finance Series (Book 396)*, Wiley, Chichester.

[53] Whitcher, B. 2004. “Wavelet-based estimation for seasonal long-memory processes.” *Technometrics* 46.

[54] Woodward, W.A., Q.C. Cheng, and H.L. Gray. 1998. “A k-factor GARMA long-memory model”. *J. Time Ser. Anal.* 19 (4): 485-504.

[55] 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.

[56] Zhang, G.P. 2003. “Time series forecasting using a hybrid ARIMA and neural network model.” *Neurocomputing* 50: 159-175.

[57] Zhang, G.P., and G.M. Qi. 2005. “Neural network forecasting for seasonal and trend time series”. *European Journal of Operational Research* 160: 501-514.

[58] Zhang, Y., Li. C., Li, L. 2017. «Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods.” *Applied Energy*, 19, 291-305