



HAL
open science

Wavelet-based multi-resolution analysis and artificial neural networks for forecasting temperature and thermal power consumption

Julien Eynard, Stéphane Grieu, Monique Polit

► **To cite this version:**

Julien Eynard, Stéphane Grieu, Monique Polit. Wavelet-based multi-resolution analysis and artificial neural networks for forecasting temperature and thermal power consumption. *Engineering Applications of Artificial Intelligence*, 2011, 24 (3), pp.501-516. 10.1016/j.engappai.2010.09.003 . hal-00567529

HAL Id: hal-00567529

<https://hal.science/hal-00567529v1>

Submitted on 21 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Wavelet-based multi-resolution analysis and artificial neural networks for forecasting temperature and thermal power consumption

Julien Eynard, Stéphane Grieu¹ and Monique Polit

*ELIAUS Lab., University of Perpignan Via Domitia, 52 Av. Paul Alduy, 66860, Perpignan, France
julien.eynard@univ-perp.fr ; grieu@univ-perp.fr ; polit@univ-perp.fr*

Abstract: as part of the OptiEnR research project, the present paper deals with outdoor temperature and thermal power consumption forecasting. This project focuses on optimizing the functioning of a multi-energy district boiler (La Rochelle, west coast of France), adding to the plant a thermal storage unit and implementing a model-based predictive controller. The proposed short-term forecast method is based on the concept of time series and uses both a wavelet-based multi-resolution analysis and multi-layer artificial neural networks. One could speak of "MRA-ANN" methodology. The discrete wavelet transform allows decomposing sequences of past data in subsequences (named coefficients) according to different frequency domains, while preserving their temporal characteristics. From these coefficients, multi-layer Perceptrons are used to estimate future subsequences of 4 hours and 30 minutes. Future values of outdoor temperature and thermal power consumption are then obtained by simply summing up the estimated coefficients. Substituting the prediction task of an original time series of high variability by the estimation of its wavelet coefficients on different levels of lower variability is the main idea of the present work. In addition, the sequences of past data are completed, for each of their components, by both the minute of the day and the day of the year to place the developed model in time. The present paper mainly focuses on the impact on forecast accuracy of various parameters, related with the discrete wavelet transform, such as both the wavelet order and the decomposition level, and the topology of the neural networks used. The number of past sequences to take into account and the chosen time step were also major concerns. The optimal configuration for the tools used leads to very good forecasting results and validates the proposed MRA-ANN methodology.

Keywords: wavelet-based multi-resolution analysis, artificial neural networks, outdoor temperature, thermal power consumption, hot water distribution network, district boiler.

1. Introduction

The actual European energy context reveals that the building sector is one of the largest sectors of energy consumption. In France, about 25% of Greenhouse Gases (GHG) emissions and 45% of energy consumption are due to buildings [1]. Consequently, the adopted "Energy Performance of Buildings Directive" (Official Journal of the European Communities, 2002) [2], focusing on energy use in buildings, requires all the European Union (EU) members to enhance their building regulations and to improve energy efficiency. With the aim of surmounting the actual energy crisis (mainly caused by both the rarefaction of fossil fuels and excessive energy consumption), managing energy demand, promoting renewable energy and finding ways for energy savings are worldwide concerns [3].

The present paper, the first of a series of three dealing with the modeling and the optimization of a district boiler as part of the OptiEnR project, focuses on the development of a forecasting tool using a wavelet-based multi-resolution analysis and artificial neural networks. The OptiEnR project began in late 2008 and will proceed, at least, until late 2010. It includes researchers from the ELIAUS laboratory of the University of Perpignan Via Domitia (south of France) and engineers from two French companies, Cofely GDF-SUEZ [4] and Weiss France [5]. Cofely GDF-SUEZ is currently the leading European brand for environmental and energy efficiency services while Weiss France designs, manufactures and installs automatic boiler rooms for all types of wood, biomass, wastes and fluids. The considered district boiler, which will be described later in the paper, is situated at La Rochelle, west coast of France, and managed by Cofely GDF-SUEZ. It supplies domestic hot water and heats residential and public buildings, using mainly wood and sometimes fuel or gas if necessary. The OptiEnR project focuses on optimizing the performance of the boiler, adding to the plant a thermal storage unit and using a model-based predictive optimal controller. Its main objective is to minimize the use of fossil energy, stocking renewable energy

¹ Corresponding author. Tel.: +33 4 68 66 22 02; Fax: +33 4 68 66 22 87.

during low-demand periods and using it during high-demand periods. The project consists of the following successive tasks (Figure 1): the first task is to forecast outdoor temperature and thermal power consumption (of the hot water distribution network); the second task focuses on modeling the district boiler; the third task deals with studying the feasibility of integrating to the plant a thermal storage unit and finding the most adequate storage material; the fourth task is to model the thermal storage while the fifth task focuses on optimizing the boiler functioning using both a Model Predictive Controller (MPC) [6-7] and the forecasted parameters (first task). As above-mentioned, the present paper deals with the first task, the next papers of the series dealing with the tasks 2 (paper 2) and 3 to 5 (paper 3).

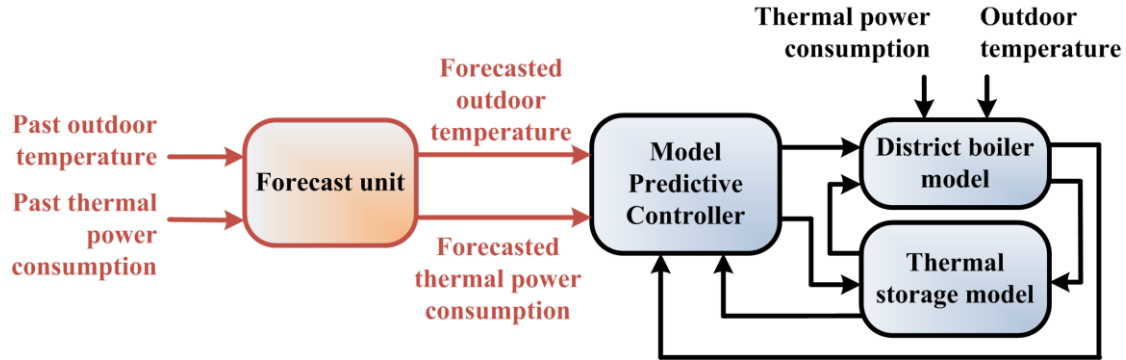


Figure 1. Diagram of the OptiEnR project. In red, the forecast module presented in the paper.

Schemes for hourly temperature forecasting have been mainly developed in the context of short or long-term load forecasting and power utilities management. The complex and nonlinear nature of temperature variations as well as the abundance of historical data suggest that computational intelligence data-based modeling techniques would be good candidates to forecast temperature. In 1996, Liu *et al.* [8] evaluated fuzzy logic, artificial neural networks and auto-regressive models for very short-term load forecasting. They concluded that the just-mentioned tools can be good candidates for this application. In 1996 and 1997, Khotanzad *et al.* [9-10] also used backpropagation neural networks to forecast hourly temperature for the next seven days, using daily high and low temperatures and an adaptive daily update of the weights. The main drawback of this approach is that the large size of the neural networks involved leads to a large number of weights to be optimized. In 1998, Hwang *et al.* [11] also used artificial neural networks and proposed a new and effective learning algorithm allowing solving both the problems of underfitting and overfitting. Yoo and Pimmel [12] developed in 1999 a self-supervised adaptive neural network to perform short term load forecasts. They used this kind of neural networks to extract correlational features from temperature and load data. The levels of error compared favorably with those given by other techniques. In 2000, Sharif and Taylor [13] used separate artificial neural networks for load forecasting of one hour to four hours ahead. We can also highlight that González and Zamarreño [14] used in 2001 a radial basis functions neural network to develop a short-term temperature forecaster, without using the daily high and low temperature values. They obtained good results. One year later (2002), the same authors, González and Zamarreño [15], proposed a short-term hourly environmental temperature forecaster based on a state space neural network. Its training was based on a random optimization method. Because of the non-stationary characteristic of temperature, training was executed daily with the aim of updating the network weights. The outdoor temperature dynamics was satisfactorily captured. However, and, as it was highlighted by Hippert *et al.* in 2000 [16], a large input dimensionality compared with the number of training records can lead to unstable networks and/or overfitting. As a consequence, one can remark that Tassadducq *et al.* [17] used in 2002 a backpropagation neural network to forecast the temperature at a given hour of the next day, only using the temperature at the same hour of the present day. In 2008, Chaabene and Ben Ammar [18] introduced a new methodology for dynamic forecasting of meteorological parameters. An ANFIS [19-21] neuro-fuzzy system offered daily time distribution of irradiance and ambient temperature relying on the meteorological behavior during the days before. First, the obtained medium term forecasting was then modeled using an ARMA model [22-23]. Next, a Kalman filter [24-26] provided short-term forecasting. The developed models were tested and validated, leading again to very good results.

Because artificial neural networks also suffer from the difficulty of finding both an optimal topology and adequate training parameters [27], some different approaches were proposed. In 1957, Allen [28] tested a simple parametric model for minimum temperature prediction. In 1985, Schneider *et al.* [29] fitted a two-harmonics Fourier model to the temperature data of the past 21 days to produce a temperature day profile. Hourly forecasts were then obtained by stretching or contracting this profile. Fan and McDonald [30] proposed in 1994 a load forecasting model consisting of time series, nonlinear load-weather functions and a residual function represented by an ARMA model [22-23]. In 1996, Chow and Leung [31] presented a novel technique for electric load forecasting based on neural weather compensation. They proposed a nonlinear generalization of the Box and Jenkins approach [32-33] for nonstationary time series forecasting and obtained very good results. In 2000, Chen and Hwang [34] used fuzzy time series [35-38] to deal with temperature forecasting problems and overcome the drawback related to historical data represented by linguistic values. They proposed a new fuzzy time series model, called two-factor time-variant fuzzy time series model. Based on the proposed model, they developed two algorithms for temperature prediction and obtained good forecasting results. Using the same model, Lee *et al.* [39] established in 2006 two-factor high-order fuzzy logical relationships based on historical data. Again, the proposed method got a higher forecasting accuracy rate than the existing methods. In 2004, Abdel-Aal [40] used abductive networks to forecast hourly temperature. Abductory inductive mechanism is a supervised inductive machine-learning tool for automatically synthesizing abductive network models from a database of inputs and outputs representing a training set of solved examples. This tool can automatically synthesize adequate models that embody the inherent structure of complex and highly nonlinear systems. Abdel-Aal developed next-day and next-hour models. Performance compares favorably with neural networks models developed using the same data, and with more complex neural networks that requires daily training. Performance is significantly superior to naïve forecasts based on persistence and climatology. In 2008, Dong and Pedrycz [41] introduced the concept of granular time series. It can be mainly applied to long-term forecasting and trend forecasting to overcome the curse of dimensionality (which plagues most predictors when carrying out long-term forecasts) and cope with uncertainly present in many times series. Dong and Pedrycz used a technique of fuzzy clustering to construct information granules on a basis of available numeric data present in the original time series. In the sequel, they developed a temperature forecasting model which captures the essential relationships between such information granulates and in this manner constructs a fundamental forecasting mechanism. Dong and Pedrycz demonstrated that the model they proposed comes with a number of advantages which manifest when processing with a large number of data. The same year (2008), Lee *et al.* [42] presented a new method for temperature prediction, based on high-order fuzzy logical relationships and genetic simulated annealing techniques [43-47]. Simulated annealing techniques were used to adjust the length of each interval in the universes of discourse with the aim of increasing the forecasting accuracy rate. The proposed method provided very good accuracy.

One can also highlight some interesting and recent works, although not directly related to temperature forecasting but having many similarities with the proposed methodology. In 2007, González-Romera *et al.* [48] used artificial neural networks to extract the trend component from monthly electric demand data and then performed separate predictions of both tendency and fluctuation, which were summed up to obtain the series forecasting. A mean absolute percentage error of about 2% was obtained. The same authors, González-Romera *et al.*, proposed a novel hybrid approach to investigate the periodic behavior of the Spanish monthly electric demand series [49]: this behavior is forecasted with a Fourier series [50] while the trend is predicted using an artificial neural network. Satisfactory results were obtained. These results improve those reached when only neural networks or ARIMA models [51-52] were used for the same purpose. As highlighted by the authors, a correct separation of trend and fluctuation and the optimization of the forecasting tools used to carry out the predictions are the key issues in the success of this technique. Furthermore, the combined use of a neural network and a Fourier series provides a simpler structure than that with only one neural network to carry out the whole prediction.

The proposed methodology deals with the concept of time series, even if not only past values are considered for estimating future values, and uses a wavelet-based Multi-Resolution Analysis and multi-layer Artificial Neural Networks. One could speak of "MRA-ANN" methodology. The discrete wavelet transform allows decomposing sequences of past data in subsequences (named coefficients) according to different frequency domains [53]. From these coefficients, Multi-Layer Perceptrons (MLP) are used to estimate future subsequences of 4 hours and 30 minutes. Future

values of outdoor temperature and thermal power consumption are then obtained by simply summing up the estimated coefficients. Substituting the prediction task of an original time series of high variability by the prediction, using MLP neural networks (many other neural networks have been tried but no significant improvement of the accuracy was observed), of its wavelet coefficients on different levels of lower variability, is the main idea of the present work. Then, the reconstruction of future values is performed by simply summing up the estimated coefficients. In addition, the sequences of past data are completed, for each of their components, by both the minute of the day and the day of the year to place the developed model in time. First, the paper focuses (section 2) on the considered district boiler and its working, and highlights the required upgrades. Section 3 describes both the data used to develop and validate the forecasting tool and the proposed MRA-ANN methodology. The wavelet-based multi-resolution analysis and the multi-layer Perceptron are also described. Sections 4 (outdoor temperature) and 5 (thermal power consumption of the hot water distribution network) present the forecasting results. Section 4 deals also with the impact on forecast accuracy of parameters such as both wavelet order and decomposition level, and with the topology of the ANN used. The final part of the paper concludes the present work and remembers the aim of the OptiEnR project.

2. The district boiler of La Rochelle

The district boiler of La Rochelle, whose synopsis is shown in Figure 2, is composed of a breaking pressure bottle, a cogeneration plant and two thermal boilers. The first one, a large 4.5 MW wood boiler, uses renewable energy (it is fed with woodchip). The second one, a 7 MW gas-fuel oil boiler, uses fossil energy. Both boilers have smokes and air-to-water heat exchangers. They supply hot water to the collecting hydraulic circuit according to a temperature set-point which is defined from outdoor temperature. During the cold season (from October to May), the wood boiler is continuously running while its heating power is adapted to the demand. The gas-fuel oil boiler functions during very cold periods only, when the wood boiler fails to respond to the demand. The primary hydraulic circuit (3000 m³), or "distribution network", supplies hot water to heat residential and public buildings (for example schools), for a total of 2700 accommodations. Domestic hot water is also produced, for a total of 3500 accommodations. The cogeneration plant, connected to the "return" part of the primary hydraulic circuit, produces electricity using gas and warms up the cold water before it goes back to the collecting hydraulic circuit. The breaking pressure bottle pulls apart the two hydraulic circuits, because of the difference between their respective flows.

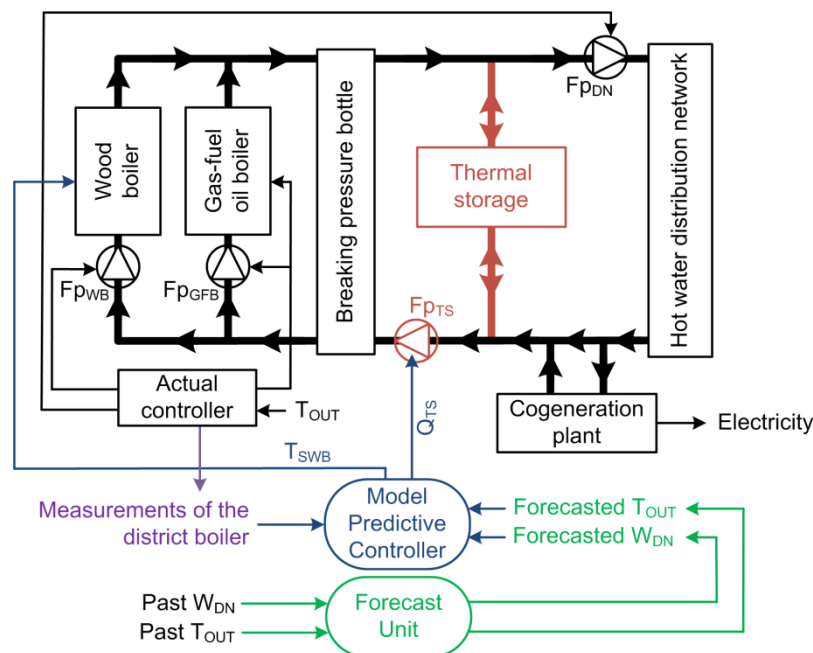


Figure 2. Synopsis of the district boiler of La Rochelle. Its current situation (in black) and future perspectives (in red, blue and green).

Table 1 presents the coverage rate of each of the three heat generators. Let us remember that the main goal of the OptiEnR project deals with the minimization of the gas-fuel oil boiler coverage rate, leading to the reduction of the fossil energy consumption. As a consequence, the wood boiler coverage rate will be maximized. To reach this objective, we proposed to add a thermal storage unit to the plant (Figure 2) with the aim of stocking hot water when the demand is low and using it when the demand can't be met by the wood boiler. This is a classical solution to optimize the functioning of boilers used to heat buildings [54-56]. Thus, the gas-fuel oil boiler will be only used when both the hot water demand is very high and the thermal storage unit is empty. The flow of the water passing through this unit (Q_{TS}) will be adjusted thanks to the control of the thermal storage feed pump (Fp_{TS}). With the aim of optimizing the use of this thermal storage unit and reducing the fossil boiler coverage rate, a model predictive controller will define over the next 4 hours and 30 minutes the optimal sequence of Q_{TS} and the wood boiler set-point temperature (T_{SWB}), taking into consideration some parameters measured at the district boiler as well as both the forecasted outdoor temperature (T_{out}) and thermal power consumption (of the distribution network) (W_{DN}). The wood boiler (Fp_{WB}) and gas-fuel oil boiler (Fp_{GFB}) feed pumps are controlled using a standard on/off controller. Fp_{DN} is the distribution network feed pump allowing controlling the water differential pressure according to outdoor temperature variations (Figure 2).

Table 1. Coverage rates of the three heat generators.

Heat generator	Coverage rate
Wood boiler	50%
Gas-fuel boiler	15 to 20%
Cogeneration plant	30 to 35%

3. Materials and methods

3.1. Database: outdoor temperature and thermal power consumption

Outdoor temperature (T_{out}) is measured at the district boiler and used for computing both the boilers and the primary circuit temperature set-points and the differential pressure set-point. That is why forecasting accurately this parameter, also used by the Model Predictive Controller (MPC) for managing the thermal storage unit, is of paramount interest. When taking a look at the data, one can highlight some interesting characteristics: first, outdoor temperature increases during the considered period (one speaks of "increase trend"); secondly, a 24-hour pseudo-period has been detected; finally, climatic phenomena impact on the parameter variability throughout the day. Of course, we want the proposed forecast methodology, based on a multi-resolution analysis and artificial neural networks, to be able of detecting these characteristics and taking advantage of them to make accurate forecasts. Let us also note that a meteorological station located in La Rochelle measures outdoor temperature and provides to the district boiler operators forecasted temperatures for five future days (from D+3 to D+7), with a sampling time set to 3 hours. With the aim of comparing these measured and forecasted temperatures with the temperatures measured at the district boiler, the considered vectors are upsampled from 3 hours to 5 minutes. Indeed, the sampling time of all the parameters measured and used to develop the boiler model is set to 5 minutes. Table 2 allows comparing both the forecasted and measured outdoor temperatures with the measurements done at the district boiler of La Rochelle.

Table 2. Comparison of both the forecasted and measured outdoor temperatures provided by the meteorological station with the measurements done at the district boiler. FIT: similarity criterion; MRE: Mean Relative Error.

Outdoor temperature	FIT (%)	MRE (%)
Measured at the meteorological station	59.7	4.17
Forecasted at the meteorological station (T_{out}^{D+3})	43.1	6.14
Forecasted at the meteorological station (T_{out}^{D+4})	38.9	6.61
Forecasted at the meteorological station (T_{out}^{D+5})	34.2	6.99
Forecasted at the meteorological station (T_{out}^{D+6})	27.6	7.88
Forecasted at the meteorological station (T_{out}^{D+7})	17.1	9.01

Table 2 highlights first that both the outdoor temperatures measured at the meteorological station and at the district boiler are quite similar. As a result, one can suppose that the temperature forecasted at the meteorological station is usable for managing the thermal storage unit using a MPC. Indeed, both the temperatures forecasted at the station and measured at the district boiler are also quite similar (Figure 3), with a FIT and a MRE ranging between 17.1% (T_{out}^{D+7}) and 43.1% (T_{out}^{D+3}) and between 6.14% (T_{out}^{D+7}) and 9.01% (T_{out}^{D+3}) respectively. Unfortunately, Figure 3 highlights that the peaks of temperature are significantly under-estimated. As a consequence, the MPC could give the order to start the gas-fuel oil boiler unnecessary. So, instead of improving the district boiler performance and decreasing the use of fossil energy, more gas and fuel will be consumed. That is why, finally, the outdoor temperature forecasted at the meteorological station will not be used by the MPC, a forecast module being developed and presented in this paper.

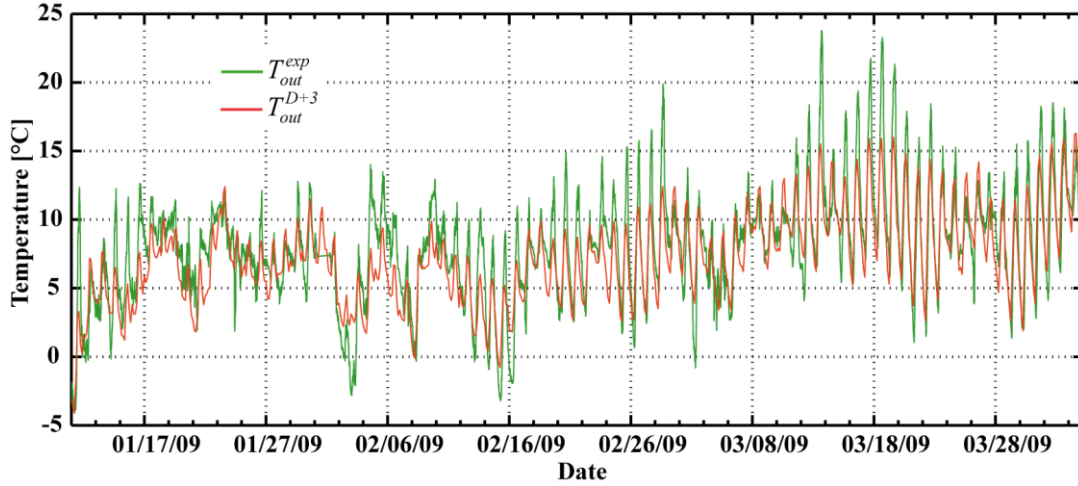


Figure 3. The outdoor temperatures forecasted at the meteorological station (T_{out}^{D+3}) and measured at the district boiler (T_{out}^{exp}) (from early January to late March).

Figure 4 depicts the calculated thermal power consumption of the hot water distribution network (W_{DN}^{exp}) during the first months of 2009 (from early January to late March). This parameter is not measured *in situ* and has an exogenous influence on the district boiler functioning because it impacts on the temperature of the water back from the distribution network. W_{DN}^{exp} is calculated using the difference between $T_{DN}^E = T_{BPP}^L$ (the temperature of the hot water leaving the breaking pressure bottle and entering the distribution network) and T_{DN}^L (the temperature of the cold water coming back from the distribution network) as well as the flow of the water passing through this network (Q_{DN}). Q_{DN} is estimated using the opening percentage of the distribution network feed pump valve ($\%_{fp}$), related to the water differential pressure (P_{DN}), its set-point (P_{SDN}) and outdoor temperature (Figure 2). As depicted by Figure 4, the thermal power consumption and the outdoor temperature are inversely proportional.

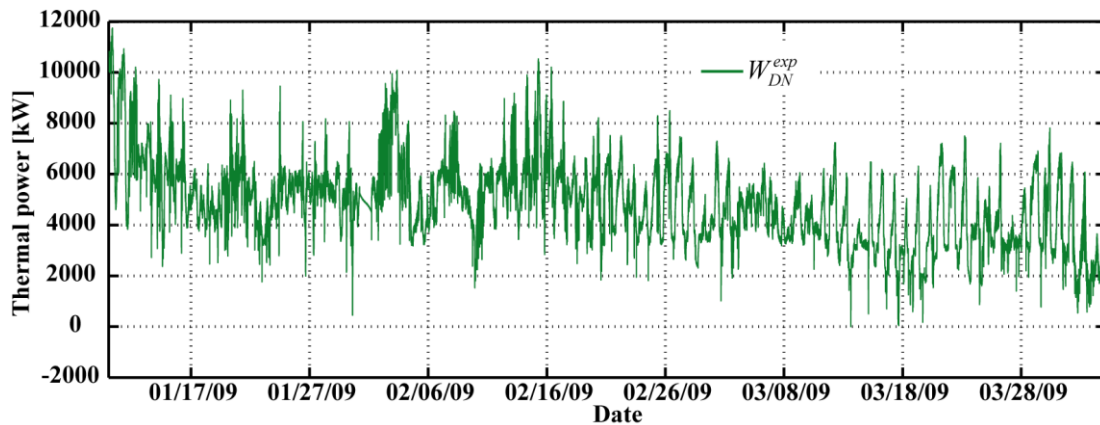


Figure 4. Calculated thermal power consumption of the hot water distribution network (W_{DN}^{exp}).

3.2. Wavelet-based multi-resolution analysis

3.2.1. Continuous Wavelet Transform (CWT)

The continuous wavelet transform is used with the aim of decomposing a signal into wavelets i.e. into highly localized small oscillations [57-58]. Whereas the Fourier transform [59-60] decomposes a given signal into infinite length sines and cosines, losing all time-localization information, the CWT's basis functions are scaled and shifted versions of the time-localized mother wavelet. So, this mother wavelet is a continuous in both time and frequency function that can be considered as a source function from which scaled and translated basis functions are generated. The CWT is used to obtain a time-frequency representation of a signal that offers very good both time and frequency localization. Considering a mother wavelet $\Psi(t)$, one can obtain the following wavelet series $\Psi_{a,b}(t)$, where a is the scale factor and b is the translation factor according to the following expression (equation 1) [61]:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right) \quad a, b \in R; a \neq 0 \quad (1)$$

For any function $x(t) \in L^2(R)$, its continuous wavelet transform can be expressed as follows (equation 2):

$$C_{a,b} = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t) \cdot \Psi^*\left(\frac{t-b}{a}\right) dt \quad (2)$$

where Ψ^* is the complex conjugate of Ψ . $C_{a,b}$ is called the wavelet coefficient. The continuous wavelet coefficients can be used with the aim of reconstructing the function $x(t)$ in the following way (equation 3):

$$x(t) = \frac{1}{C_\Psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} C_{a,b} \Psi_{a,b}(t) \frac{da db}{a^2} \quad (3)$$

3.2.2. Discrete Wavelet Transform (DWT) and bank of filters

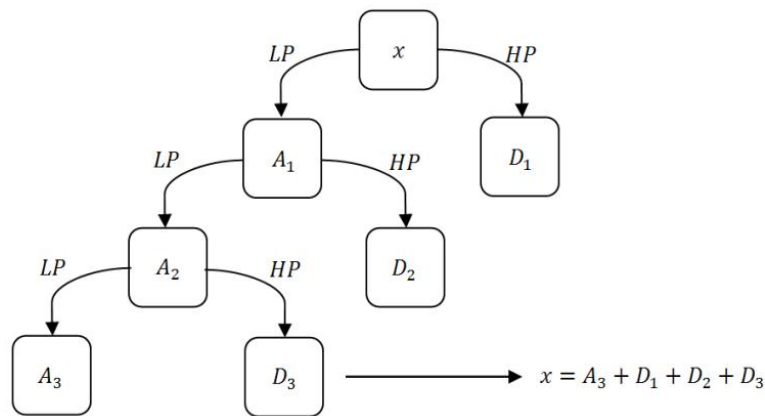


Figure 5. Multi resolution analysis leading to the 3-level decomposition of a signal x .

Because translating the mother wavelet leads to redundant information, one can use, instead of the continuous wavelet transform, the discrete one. In this case, both scale (a) and translation (b) factors are restricted to only discrete values such that $a = a_0^m$ and $b = nb_0 a_0^m$, where $a_0 > 1$, $b_0 \neq 0$, and $(m, n) \in Z^2$. Basically, the discrete wavelet multi-resolution analysis, commonly based on Daubechies orthogonal wavelet basis [57], allows decomposing a signal into approximations (i.e. low frequency coefficients) and details (i.e. high frequency coefficients) using a filter bank composed of both low-pass (LP) and high-pass (HP) filters [62]. This process can be repeated n times, producing n levels of decomposition, but decomposing (downsampling) the approximations

(the low frequency coefficients) only. Indeed, the high frequency coefficients are neglected. So, a signal x can be first decomposed into an approximation A_1 and a detail D_1 (that is the level 1 of the decomposition). Then A_1 can be decomposed into an approximation A_2 and a detail D_2 (that is the level 2 of the decomposition) and so on. Considering n levels of decomposition, the reconstruction process allows recovering the initial signal, summing the n details D_1, D_2, \dots, D_n and the approximation A_n of level n . As an example, Figure 5 depicts the 3-level decomposition of a signal x and its reconstruction [62].

3.2.3. The Daubechies wavelets family

Different families of wavelets whose qualities vary according to several criteria can be used for analyzing sequences of data points [63]. The main criteria are (1) the speed of convergence to 0 of these functions when the time t or the frequency w goes to infinity, which quantifies both time and frequency localizations, (2) the symmetry, (3) the number of vanishing moments of Ψ and (4) the regularity, which is useful for getting nice features, like smoothness of the reconstructed signal. The most commonly used wavelets are the orthogonal ones (Daubechies, Symlet or Coiflet wavelets) [64]. Because the Daubechies wavelets [57] (Figure 6) have the highest number of vanishing moments, this family has been chosen for carrying out the wavelet-based multi-resolution analysis of the considered sequences of data points.

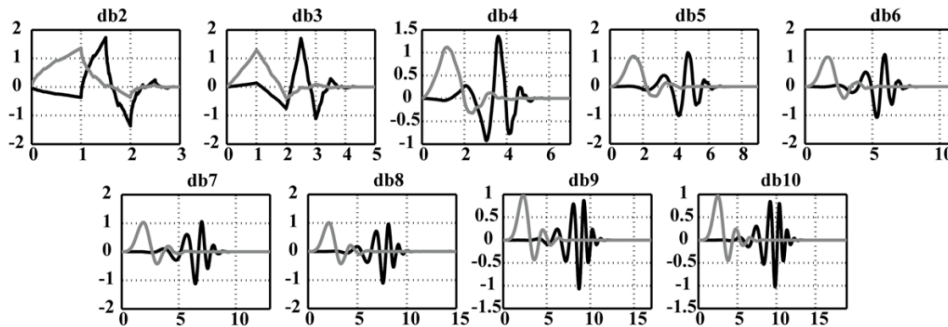


Figure 6. The Daubechies wavelets family, according to the wavelets order. In black: the wavelets, in grey: the associated scaling functions.

3.3. The multi-layer Perceptron neural network

3.3.1. Network topology

The Perceptron, the simplest neural network, is only able to classify data into two classes [65]. Basically it consists of a single neuron with a number of adjustable weights [66]. It uses an adaptative learning rule. Given a problem which calls for more than two classes, several Perceptrons can be combined: the simplest form of a layered network just has an input layer of source nodes that connect to an output layer of neurons. The single-layer Perceptron can only classify linearly separable problems. For non-separable problems it is necessary to use more layers. A multi-layer network has one or more hidden layers whose neurons are called hidden neurons. The network is fully connected, every node in a layer is connected to all nodes in the next layer. According to the previous remarks, the network used for the present work is a multi-layer Perceptron. It consists of one layer of linear output neurons and one hidden layer of nonlinear neurons [67] (Figure 7). According to previous tests, more than one hidden layer proved to cause slower convergence during the learning phase because intermediate neurons not directly connected to output neurons learn very slowly. Based on the principle of generalization versus convergence, both number of hidden neurons and iterations completed during the training phase were optimized [68]. The multi-layer Perceptron neural network learns using an algorithm called backpropagation. During this iterative process, input data are repeatedly presented to the network. With each presentation, the network output is compared to the desired output and an error is computed. This error is then fed back to the network and used to adjust the weights such that it decreases with each iteration and the model gets closer and closer to produce the desired output [69].

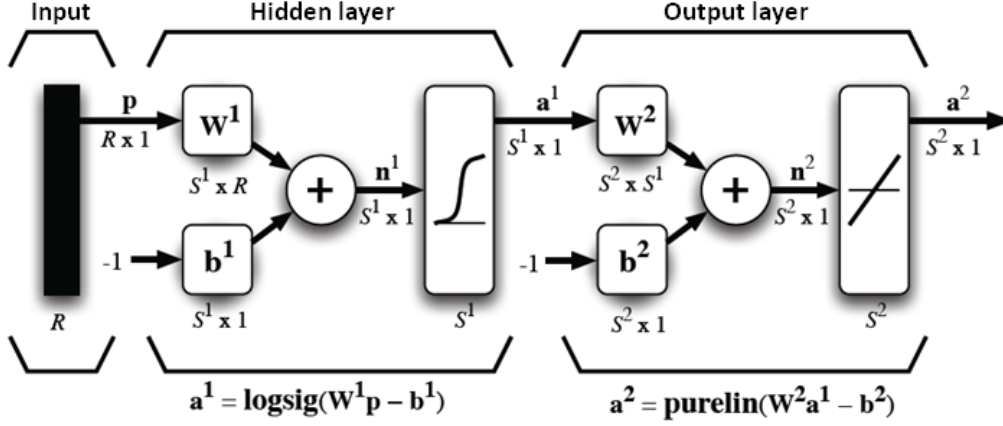


Figure 7. Topology of the multi-layer Perceptrons used for estimating the wavelet coefficients of forecasted sequences: one hidden layer of nonlinear neurons and an output layer of linear neurons.

3.3.2. The Levenberg-Marquardt algorithm

Several training methods were used, but the Levenberg-Marquardt algorithm [70] proved to be the fastest and the most robust. It is particularly adapted for networks of moderate size and has memory reduction feature for use when the training set is large. Like the quasi-Newton methods, the Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares, then the Hessian matrix can be approximated as:

$$H = J^T \cdot J \quad (4)$$

The gradient can be computed as:

$$g = J^T \cdot e \quad (5)$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. The Jacobian matrix can be computed through a standard backpropagation technique that is much less complex than computing the Hessian matrix. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$x_{k+1} = x_k - [J^T \cdot J + \mu \cdot I]^{-1} \cdot J^T \cdot e \quad (6)$$

When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift towards Newton's method as quickly as possible. Thus, μ is decreased after each successful step and is increased only when a tentative step would increase the performance function. In this way, the performance function will always be reduced at each iteration of the algorithm. The main drawback of the Levenberg-Marquardt algorithm is that it requires the storage of some matrices that can be quite large for certain problems. The size of the Jacobian matrix is $Q \times n$, where Q is the number of training sets and n is the number of weights and biases in the network. It turns out that this matrix does not have to be computed and stored as a whole. For example, if we were to divide the Jacobian into two equal submatrices we could compute the approximate Hessian matrix as follows:

$$H = J^T \cdot J = [J_1^T \quad J_2^T] \cdot \begin{bmatrix} J_1 \\ J_2 \end{bmatrix} = J_1^T \cdot J_1 + J_2^T \cdot J_2 \quad (7)$$

Therefore, the full Jacobian does not have to exist at one time. The approximate Hessian can be computed by summing a series of subterms. Once one subterm has been computed, the corresponding submatrix of the Jacobian can be cleared.

3.4. Forecast methodology

This section deals with the description of the proposed forecast methodology, based on both the concept of time series and the use of a wavelet-based multi-resolution analysis and multi-layer artificial neural networks. One speaks of MRA-ANN methodology.

3.4.1. Overall approach

A time series is a sequence of data points, measured typically at successive times and spaced at, often uniform, time intervals [71]. Time series forecasting is the use of a model to forecast future events based on known past events [72]. The aim of the proposed MRA-ANN methodology is forecasting, thanks to artificial neural networks, a sequence of l data points, using M ($M \geq 1$) past sequences, also of length l . According to the respective values of M , l and of the sampling time T (set initially to 30 minutes; forecasted sequences are then upsampled to be in agreement with the sampling time of all the parameters measured at the district boiler), the results can be more or less accurate. Indeed, the further the forecasting horizon (set to 4h30, as requested by the boiler operators and taking into account the main objective of the OptiEnR research project, this leads to $l = 9$), the more inaccurate it can be. Inaccuracy can also be the result of a lack of usable information provided by past sequences or the result of a too low number M of considered past sequences. Past sequences are examples to be learned and used to forecast future values of the considered parameters (in this case, the outdoor temperature and the thermal power consumption of the hot water distribution network). To help the model to place itself in time, past sequences are completed, for each of their components, by the minute of the day and the day of the year. The proposed neuronal model has been developed (artificial neural networks have been trained) using half of the available sequences and validated thanks to the remaining sequences (Figure 8).

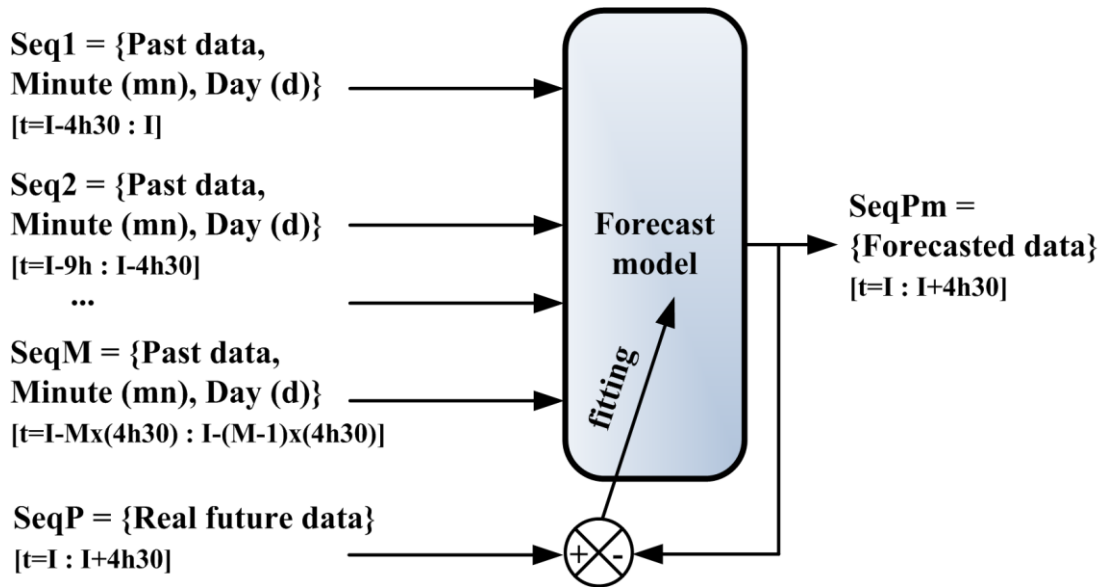


Figure 8. Development phase of the forecast neuronal model.

3.4.2. Wavelet-based multi-resolution analysis contribution

Since the data usually consist of low and high frequencies components including noise, obtaining the essential frequency-domain features from these data becomes an important aspect of data analysis, visualization and forecasting. As a consequence, understanding and being able to learn from data, using artificial neural networks, is not as easy as one may think. Usually, tools such as artificial neural networks learn very well, optimizing their topologies and all the training parameters, from limited frequency domain data. When the data consist of low and high frequencies components, a useful approach can be to develop one forecast model for each of the frequency domains and then to combine the information provided by all of the models to build the desired signal. That is why one needs a tool allowing, on one hand, decomposing a signal according to

different scales of frequency, while preserving its temporal characteristics (which is fundamental when proposing a forecast methodology based on time series), and, on the other hand, rebuilding this signal. This tool is the wavelet-based multi-resolution analysis. It allows decomposing, according to both a decomposition level (N) and a wavelet order (R), a signal into approximations and details, using a filter bank composed of low-pass and high-pass filters. Keeping in mind the previously-mentioned characteristics of both the outdoor temperature and the thermal power consumption of the hot water distribution network, one can easily understand that the wavelet-based multi-resolution analysis allows isolating the overall trend and the 24-hour pseudo period characterizing the considered time series (approximation coefficients) from the variability caused by climatic phenomena (detail coefficients). As previously-mentioned, the wavelet coefficients are of lower variability than the original and to-be-forecasted time series, which favors the estimation of future values. All the series of M past sequences (of length l) considered to develop the model are so decomposed into N details and an approximation of level N . For each series, the target sequence P is also decomposed in the same way (Figure 9).

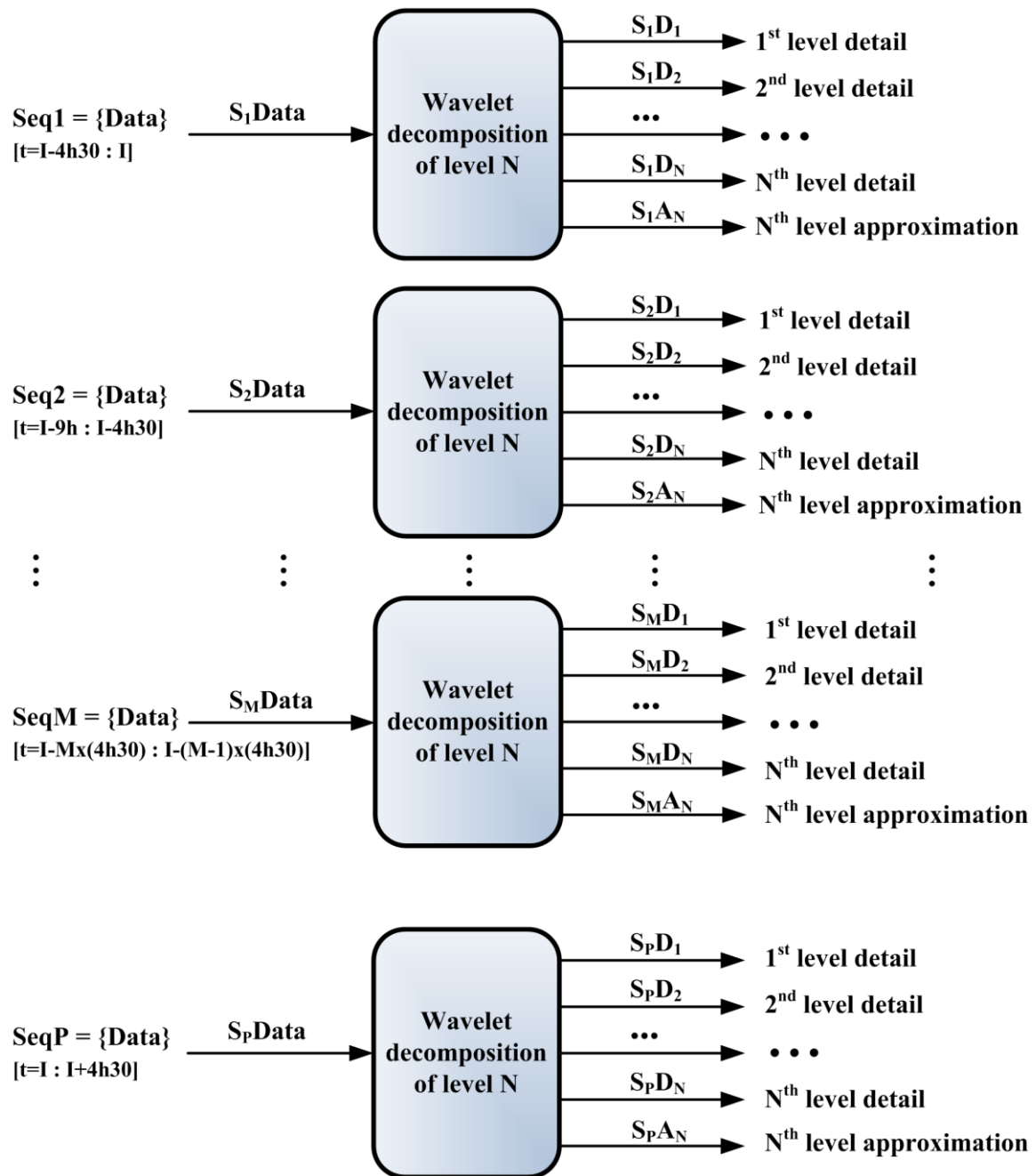


Figure 9. Decomposition of past and target sequences into N details and an approximation of level N .

3.4.3. Multi-layer Perceptrons for forecasting details and approximations of future sequences

Substituting the prediction task of an original time series of high variability by the prediction of its detail and approximation coefficients on different levels of lower variability is the key-point of the proposed MRA-ANN methodology. In this sense, a multi-layer Perceptron neural network is needed, and it has of course to be trained using the coefficients of the considered past sequences, for estimating each of the N details and the approximation of level N of the target sequence (the sequence to be forecasted). Let us remember that a sequence is composed of data points as well as temporal indicators (for each observation, the minute of the day and the day of the year). So, for a decomposition of level N , the overall forecast model is composed of $N + 1$ multi-layer Perceptrons. As previously mentioned, summing up the estimated coefficients allows rebuilding the to-be-forecasted sequence (Figure 10). Similar approaches have already been applied to the forecasting of parameters such as electric consumption [73-74], solar irradiation [75-76] or wind speed [77]. The particularities of the proposed methodology lies in the use of sequences (to train the artificial networks or to be forecasted), in the addition of temporal information allowing a better understanding of how the considered parameters (T_{out} and W_{DN}) evolve in time, and in the use of a specific artificial neural network for estimating one of the wavelet coefficients of the future values we want to forecast. Of course, and this is the main weakness of such an approach, we have to train and find the right topology for each of the neural networks used.

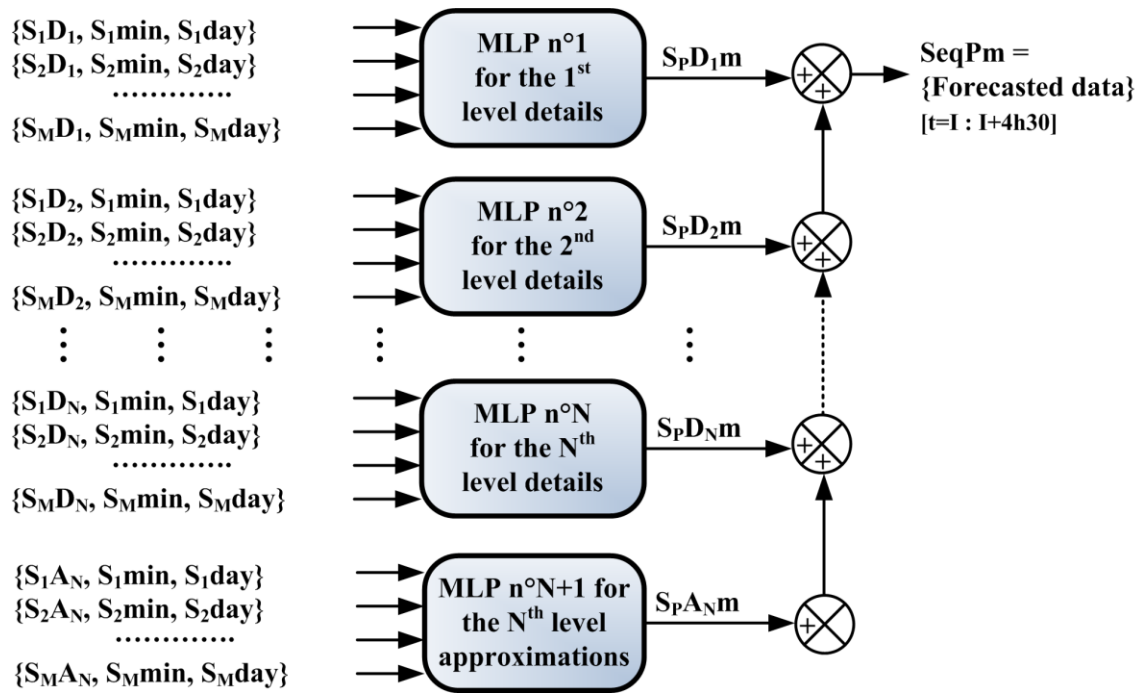


Figure 10. Multi-layer Perceptrons for forecasting details and approximations of future sequences. Rebuilding of these sequences (MRA-ANN methodology).

3.4.4. Selection of the most accurate models

As previously mentioned, the respective values of M (the number of considered past sequences), l (the length of the sequences), T (the sampling time), N (the wavelet decomposition level) and R (the wavelet order) impact the forecast accuracy. The networks topology, basically the number of hidden neurons, as well as training parameters, such as the number of iterations, the learning rate and the error goal, also impact the networks generalization capability. That is why the present paper deals with the study and the quantification of all of these influences. It focuses also on the impact on the model performance of the random initialization of all the networks synaptic weights and on the relevance of developing "specialized" models (each model is developed to forecast outdoor temperature during only a specific period of the year). As it will be highlighted in the results section and because the Mean Relative Error (MRE) distribution (considering the networks validation phase) can be considered as a Gaussian function, with more or less skewness,

one can easily select the most accurate models. According to statistical considerations, a model is selected, for a given configuration related with all the above-mentioned parameters and with a random initialization of all the network synaptic weights, when the validation error is inferior to the mean error (considering all the tested configurations) minus twice the deviation standard. The $N + 1$ selected models define the overall forecast model.

4. Outdoor temperature forecasting

This section of the paper presents experimental results in forecasting outdoor temperature. All the parameters mentioned in section 3.4.4 have been optimized to obtain the most accurate models. With the aim of reducing the complexity of this optimization problem, all the considered parameters have not been studied at the same time. As a consequence, the models configurations obtained are probably not optimal configurations but, most certainly, sub-optimal configurations. According to the common topology (the number of hidden neurons) of the multi-layer Perceptrons used, the influence of both the wavelet order and the wavelet decomposition level is, first, studied (section 4.1). Next, the impact of the number of both the considered past sequences and specialized models (section 4.2) as well as of the sampling time (section 4.3), considering optimal values of R , N and F , is studied. As mentioned in section 3.4.1, the forecasting horizon has been set to 4h30. It is a good compromise between accuracy and computation time, keeping in mind that forecasted sequences will be used by a model predictive controller to manage a thermal storage unit.

4.1. Influence of both the wavelet order and the wavelet decomposition level

As just above-mentioned, the two first parameters studied are both the wavelet order and the wavelet decomposition level. Only the number of hidden neurons varies, the sampling time is set to 30 minutes while only one past sequence of 9 points is considered ($M = 1$ and $l = 9$). Figure 11 and 12 present the validation MRE according to R , N and F (the networks common number of hidden neurons), ranging respectively between 2 and 10, between 1 and 9 and between 2 and 50. Let us note that the wavelet decomposition level has been averaged over its variation interval to obtain Figure 10. It is interesting to note that the best results (the most accurate forecasts) are achieved for a wavelet order equal to 4, using 5 hidden neurons (Figure 11). This can be explained by both the specific shape and frequency spectrum of a wavelet. This spectrum can be found, or not, in the analyzed signal. Considering, for example, a two-order wavelet, its shape is too different from the signal form and this leads to bad results.

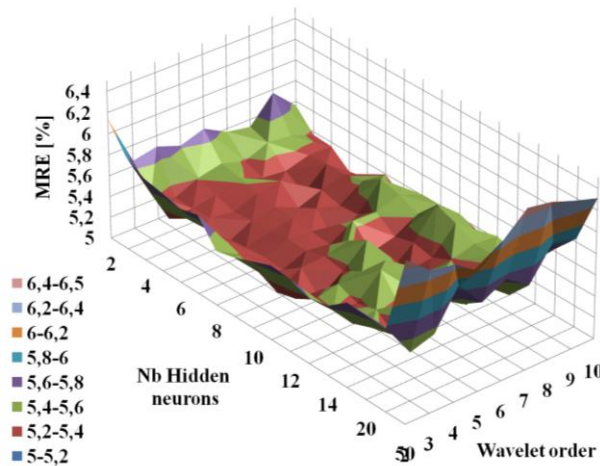


Figure 11. Influence of the wavelet order according to the networks topology.

Having a quick look at Figure 12 (the wavelet order has been averaged over its variation interval), it can be noticed that when the wavelet decomposition level is high, only few hidden neurons are needed for accurately forecasting outdoor temperature. More hidden neurons are needed when the decomposition level is low. One can consider, and this is a very interesting result, that the number of hidden neurons to be used for accurately forecasting outdoor temperature is inversely

proportional to the decomposition level. Indeed, for a too low decomposition level, each coefficient contains an important part of the initial information and, as a consequence, many hidden neurons are required to correctly model these coefficients using artificial neural networks. For a high decomposition level, the detail and approximation coefficients contain a small amount of information; that is why both their modeling and forecasting are easier and require, for each artificial neural network used, fewer hidden neurons. On the other side, let us remember that the higher the decomposition level, the more the networks needed (one for each coefficient to be estimated). The most accurate forecasts are achieved for a 5-level decomposition and 5 hidden neurons.

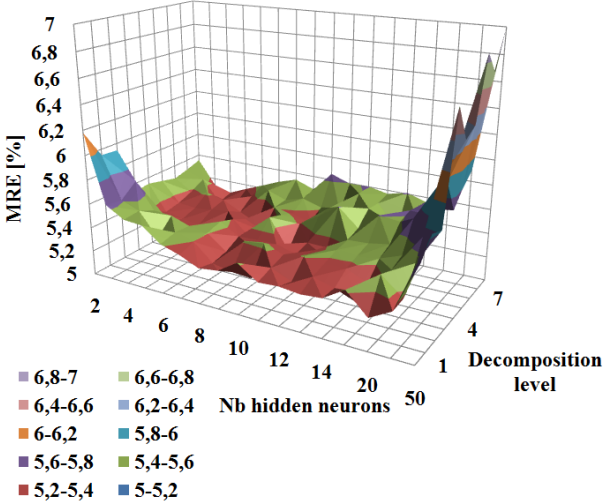


Figure 12. Influence of the wavelet decomposition level according to the networks topology (the number of hidden neurons).

Figure 13 depicts the linked influence of the wavelet order and the wavelet decomposition level (the common number of hidden neurons for all of the neural networks used has been averaged over its variation interval). It highlights specific domains of low MRE, one can speak of "valleys", and specific domains of higher MRE. Taking a look at Figure 13, it becomes apparent that, beyond an adequate wavelet decomposition level, the wavelet order is more influential than the decomposition level itself. Two valleys are clearly visible when the wavelet order is set to 4 or when it is set to 7 (the best choice seems to be 4). These valleys are very slightly dependent on the decomposition level.

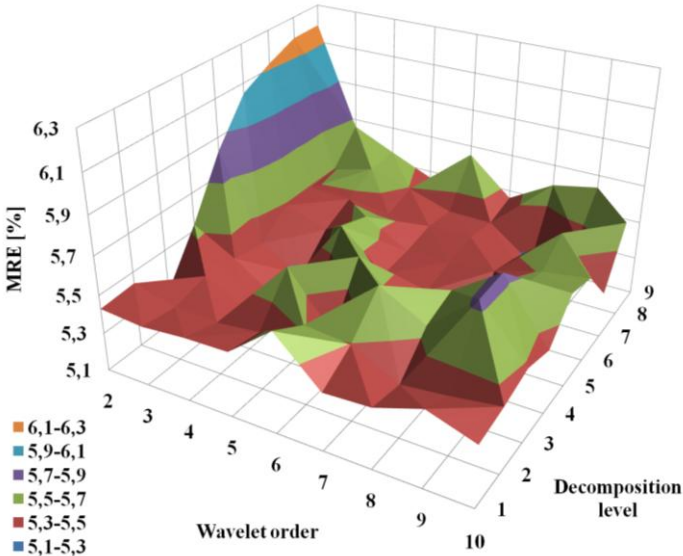


Figure 13. Linked influence of the wavelet order and the wavelet decomposition level.

As a result of this first part of the study, one can conclude that the most accurate forecasts are achieved when decomposing the considered sequences into 5 levels, using Daubechies wavelets of order 4 and multi-layer Perceptrons which respective hidden layers are composed of 5 neurons. These results were taken into account to study the influence of the number of considered past sequences. Let us note that similar results were obtained with a sampling time set to 10 minutes.

4.2. Influence of both the number of considered past sequences and the number of specialized models

In this section, the influence of both the number M of considered past sequences and the number S of specialized models is analyzed. Several simulations have been carried out, considering a number of past sequences, a number of specialized models and a number of hidden neurons ranging respectively between 1 and 8, 1 and 3 and 3 and 21. Figure 14 depicts the influence of the number of considered past sequences, considering only one overall model, according to the common topology of the neural networks used to model all of the wavelet decomposition coefficients. The results highlight the necessity of considering at least 4 past sequences (the optimal number) to obtain good forecasts. This clearly reduces the validation MRE observed. However, for a too high number of considered past sequences, the forecasts accuracy deteriorates gradually. Again, the common optimal number of hidden neurons for all of the multi-layer Perceptrons used is equal to 5. With this topology, considering 4 past sequences instead of 1 for developing and validating the overall model, the MRE is reduced of about 12.5% (from 4.8% to 4.2%).

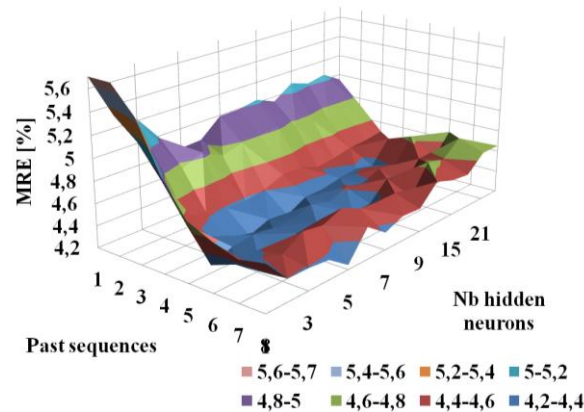


Figure 14. Influence of the number of considered past sequences according to the networks topology (one overall model).

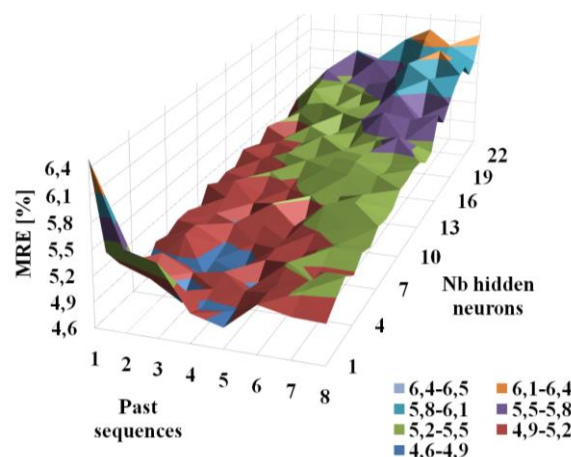


Figure 15. Influence of the number of considered past sequences according to the networks topology (three specialized models).

Figure 15 depicts the influence of the number of considered past sequences, but now considering three specialized models (each model is developed to forecast outdoor temperature during a specific period of the year only), according to the common topology of the neural networks used to

model all of the wavelet decomposition coefficients. Each of the three specialized models being trained using less (three times less) information than when developing and validating an overall model, less hidden neurons are needed for accurately estimating, using multi-layer neural networks, approximation and detail coefficients. As a consequence, the influence of a bad networks sizing (too many hidden neurons favors overfitting) is more visible.

Considering the right number of hidden neurons for all of the neural networks used to model the wavelet decomposition coefficients, Figure 16 depicts the influence of the number of considered past sequences according to the number of specialized models. Analyzing this figure, one can easily conclude that using several specialized models, instead of a unique overall model, does not improve the forecasts accuracy, quite the opposite... Whereas the main objective of developing specialized models was to dissociate and better understand (learn) the various behaviors outdoor temperature might have, one can notice that both the too restrictive historical data considered (from mid January to early April) and the networks training using less examples do not allow reducing the MRE. The networks generalization capability seems to be negatively impacted. So, one can suppose that considering, for example, one year or historical data, developing specialized models according to the various seasons of the year (temperature evolution is quite different from a season to another) would be very useful.

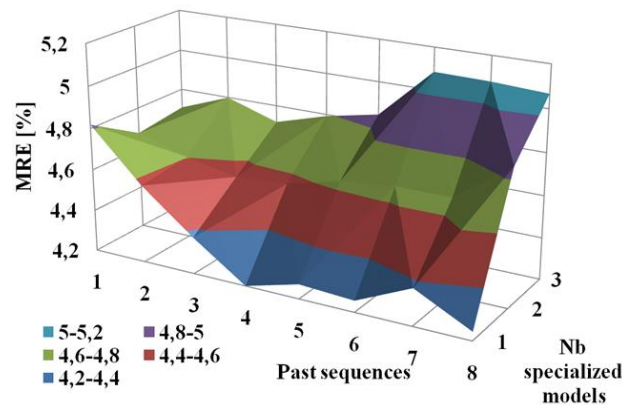


Figure 16. Influence of the number of considered past sequences according to the number of specialized models.

4.3. Influence of the sampling time

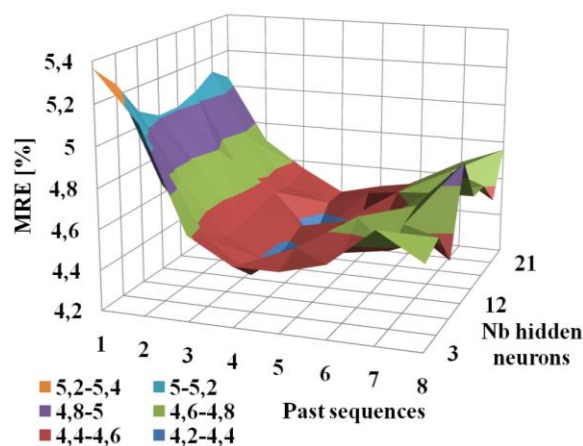


Figure 17. Influence of the number of considered past sequences according to the networks topology (sampling time = 10 minutes).

Finally, the influence of the sampling time on the models performance has also been studied, considering a wavelet order and a wavelet decomposition level equal to 4 and 5 respectively. One can conclude that reducing the sampling time from 30 minutes to 10 minutes does not impact

significantly the forecasts accuracy. Figures 17 and 18 depict, with a sampling time set to 10 minutes, the influence of the number of considered past sequences according to the topology of the neural networks used and the number of specialized models (in this case, the right number of hidden neurons for all of the neural networks used to model the wavelet decomposition coefficients is used) respectively. Both figures highlight that four past sequences and a unique overall model lead to the best results (in this case, the MRE observed is about 4.3%). A reduced sampling time impacts significantly on the common topology of the neural networks used to model all of the wavelet decomposition coefficients only. The common number of hidden neurons is about 15 (it was only of 5 with a sampling time set to 30 minutes). One can suppose that, because a reduced sampling time increases the length of the sequences (in this case, up to 27 points), more information has to be learned and, as a consequence, requires more hidden neurons for all of the neural networks used.

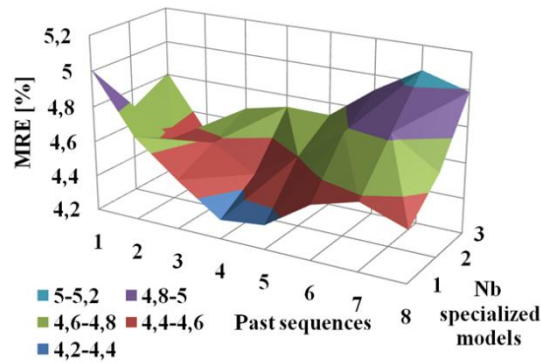


Figure 18. Influence of the number of considered past sequences according to the number of specialized models (sampling time = 10 minutes).

4.4. Optimal configuration and experimental results

As a conclusion of the study about outdoor temperature forecasting, Table 3 depicts the optimal configuration to meet performance needs when applying the MRA-ANN methodology, according to the wavelet order (R), the wavelet decomposition level (N), the topology (the common number of hidden neurons F) of the artificial neural networks used for estimating the detail and approximation coefficients, the number of considered past sequences (M), the number of specialized models (S) and, finally, the sampling time (T).

Table 3. Optimal configuration for outdoor temperature forecasting (MRA-ANN methodology).

<i>Parameter</i>	<i>Symbol</i>	<i>Optimal value</i>
<i>Wavelet order</i>	R	4
<i>Wavelet decomposition level</i>	N	5
<i>Common number of hidden neurons</i>	F	5
<i>Number of considered past sequences</i>	M	4
<i>Number of specialized models</i>	S	1
<i>Sampling time (minutes)</i>	T	30

Figure 19 allows comparing, from February 15, 2009 to March 02, 2009, the outdoor temperatures measured at the district boiler (T_{out}^{exp}), forecasted at the meteorological station (T_{out}^{D+3}) and forecasted thanks to the MRA-ANN methodology based on a multi-resolution analysis and the use of artificial neural networks (T_{out}^{4h30}). Clearly, the proposed methodology provides better forecasts than the meteorological station. The temperature peaks are far better modeled (as previously mentioned, it was one of the major objectives of the work, the forecasted temperature being used by a MPC for optimizing the boiler functioning). Indeed, one can note that frequent differences of about 5°C between the measured and forecasted (at the meteorological station) outdoor temperatures are reduced, using the developed model, to only 1°C.

Table 4 specifies, for the considered period (from February 15, 2009 to March 02, 2009) and the two forecasted outdoor temperatures (by the meteorological station and using the MRA-ANN methodology), the Mean Relative Error (MRE), the Mean Absolute Error (MAE) as well as the curve fitting (FIT) obtained. First, one can remark, considering the temperature forecasted at the meteorological station as a reference, that the developed model allows reducing both the MRE and the MAE by 33% (from 6.28% to 4.14%) and 29.4% (from 1.75°C to 1.15°C) respectively. The curve fitting is improved of about 40% (from 42.6% to 60.6%). As a conclusion, one can highlight the model performance and the validity of the proposed MRA-ANN methodology.

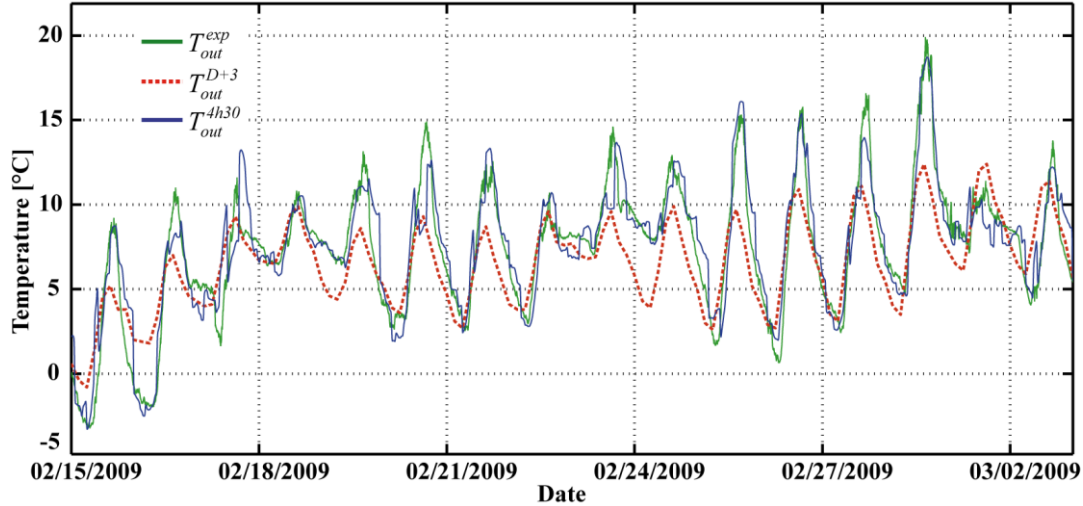


Figure 19. Outdoor temperature measured at the district boiler (T_{out}^{exp}), forecasted at the meteorological station (T_{out}^{D+3}) and forecasted using the MRA-ANN methodology (T_{out}^{4h30}).

Table 4. Forecasting performance criteria (outdoor temperature).

Forecasted outdoor temperature	MRE [%]	MAE [°C]	FIT [%]
T_{out}^{D+3}	6.28	1.75	42.6
T_{out}^{4h30}	4.14	1.15	60.6

5. Thermal power consumption forecasting

This section of the paper presents experimental results in forecasting thermal power consumption. Two different ways of doing were considered: first, using the previously forecasted outdoor temperature (both parameters are strongly correlated) and, secondly, as we did for outdoor temperature, applying the MRA-ANN methodology. Finally, a hybrid approach is proposed.

5.1. Use of the forecasted outdoor temperature

As previously mentioned (section 3.1), outdoor temperature and thermal power consumption are inversely proportional. That is why the first of the two proposed approaches deals with a simple linear curve fitting, optimizing both the parameters α and β of Equation 8 thanks to the minimization of the difference between W_{DN}^{exp} (the calculated thermal power consumption) and $W_{DN}^{T_{out}} = W_{DN}^{T_{out}^{D+3}}$ (the thermal power consumed by the hot water distribution network and estimated using the outdoor temperature forecasted at the meteorological station ($T_{out} = T_{out}^{D+3}$)) or $W_{DN}^{T_{out}} = W_{DN}^{T_{out}^{4h30}}$ (the thermal power consumed by the hot water distribution network and estimated using the outdoor temperature forecasted thanks to the MRA-ANN methodology ($T_{out} = T_{out}^{4h30}$)). Let us note that $\alpha = -381.4$ and $\beta = 7607$ if T_{out}^{D+3} is used while $\alpha = -329.6$ and $\beta = 7472.7$ with T_{out}^{4h30} .

$$\begin{cases} \min_{\alpha, \beta} \|W_{DN}^{exp} - W_{DN}^{T_{out}}\|^2 \\ W_{DN}^{T_{out}} = \alpha \cdot T_{out} + \beta \end{cases} \quad (8)$$

Table 5 (lines 1 to 2) and Figure 20 present the results of the thermal power consumption (W_{DN}) forecasting. These results validate the first proposed approach: one can easily predict the thermal power consumption of the hot water distribution network using outdoor temperature. Whatever the forecasted temperature used, the results are quite similar. For the previously-mentioned period (from February 15, 2009 to March 02, 2009), the MRE, the MAE and the FIT are about 6.97%, 761.9 kW and 37.6% when using T_{out}^{D+3} and about 6.94%, 758.3 kW and 38.4% with T_{out}^{4h30} . However, having a quick look at Figure 20, one can remark, first, that the amplitude is not always well estimated and, secondly, that the curves are not always completely in phase. So, it could be interesting to use the proposed MRA-ANN methodology for directly forecasting the thermal power consumption of the hot water distribution network.

Table 5. Forecasting performance criteria (thermal power consumption).

Thermal power consumption (W_{DN}) forecasted using:	MRE [%]	MAE [kW]	FIT [%]
T_{out}^{D+3}	6.97	761.9	37.6
T_{out}^{4h30}	6.94	758.3	38.4
MRA-ANN methodology	6.07	663.6	44.7
Hybrid approach	5.85	639.1	46.9

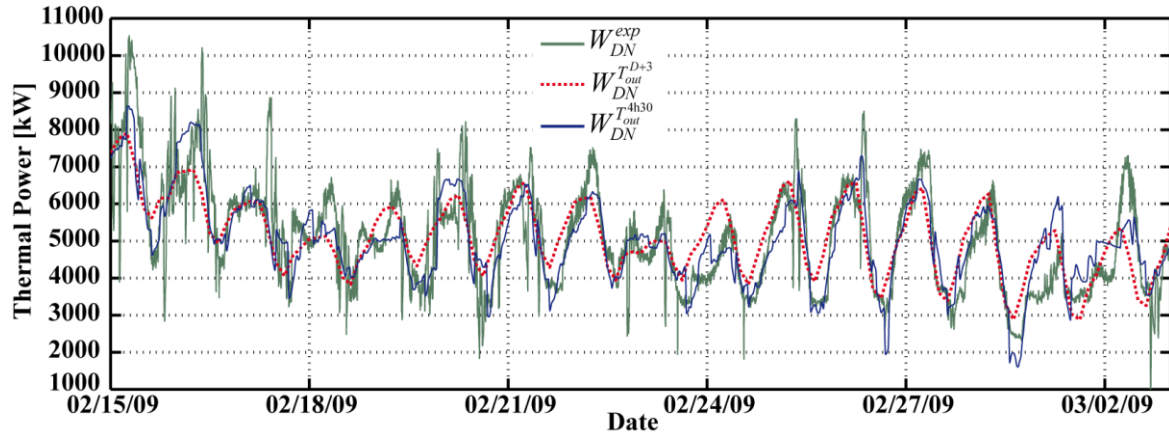


Figure 20. Thermal power consumption forecasted using T_{out}^{D+3} ($W_{DN}^{T_{out}^{D+3}}$) or T_{out}^{4h30} ($W_{DN}^{T_{out}^{4h30}}$). Comparison with the calculated W_{DN}^{exp} .

5.2. Use of the MRA-ANN methodology

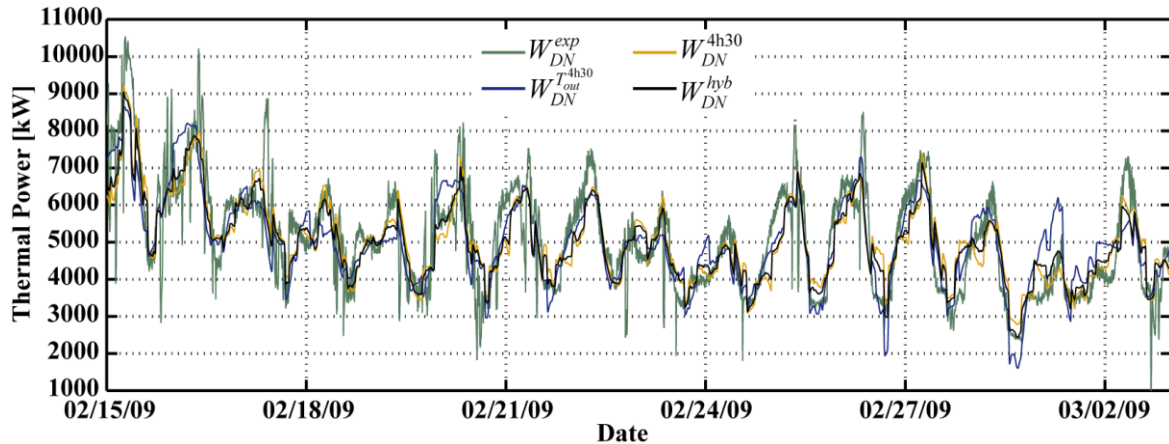


Figure 21. Thermal power consumption forecasted using T_{out}^{4h30} ($W_{DN}^{T_{out}^{4h30}}$), the MRA-ANN methodology (W_{DN}^{4h30}) and the hybrid approach (W_{DN}^{hyb}). Comparison with the calculated W_{DN}^{exp} .

The MRA-ANN methodology proposed to forecast outdoor temperature, based on a multi-resolution analysis and artificial neural networks, has also been used for directly estimating the thermal power consumption of the hot water distribution network (W_{DN}). Due to the high correlation between the just-mentioned parameters, the optimal configuration found for outdoor temperature forecasting (Table 3) has been re-used to develop and validate the new model. Table 5 (line 3) and Figure 21 present the results of the thermal power consumption forecasting, using the MRA-ANN methodology. The MRE, the MAE and the FIT are about 6.07%, 663.6 kW and 44.7% respectively. These results are better than when dealing with a linear curve fitting, using the forecasted outdoor temperature (T_{out}^{4h30}): one can highlight that the model allows reducing both the MRE and the MAE by 12.5%, from 6.94% to 6.07% and from 758.3 kW to 663.6 kW respectively. The curve fitting is improved of about 16.4% (from 38.4% to 44.7%). However, taking a look at Figure 21, one can remark that the peaks of consumption are again not always well estimated. That is why a last approach is proposed, combining time series forecasting (the MRA-ANN methodology) and linear curve fitting.

5.3. Hybrid approach

As just mentioned, a last approach is proposed, combining the MRA-ANN methodology and linear curve fitting, for predicting the thermal power consumption of the hot water distribution network. One can highlight that time series forecasting (section 5.2) provides better results than linear curve fitting (using the forecasted outdoor temperature, section 5.1) but sometimes underestimates peaks of consumption. Both approaches are valid but have different characteristics. That is why it appears that they can be used in tandem to improve the forecasts accuracy. The idea is so to combine the forecasted sequences provided by each of the two considered approaches (W_{DN}^{4h30} and $W_{DN}^{T_{out}^{4h30}}$) using a weighted mean whose coefficients γ and δ are optimized in the following way (equation 9):

$$\left\{ \begin{array}{l} \min_{\gamma, \delta} |W_{DN}^{exp} - W_{DN}^{hyb}| \\ W_{DN}^{hyb} = \gamma \cdot W_{DN}^{T_{out}^{4h30}} + \delta \cdot W_{DN}^{4h30} \end{array} \right. \quad (9)$$

The optimal values of γ and δ are about 0.28 and 0.71 respectively. As it was expected, the weight of the most accurate forecasted sequence, provided by the MRA-ANN methodology, is the highest weight. However, one can note that the influence (weight) of the sequence estimated by the "linear curve fitting" method, using the forecasted outdoor temperature, is regardless significant. Table 5 (line 4) and Figure 21 present the results of the thermal power consumption forecasting, using the just-described hybrid approach. The MRE, the MAE and the FIT are about 5.85%, 639.1 kW and 46.9% respectively. Taking as a reference the performance of the MRA-ANN methodology, one can observe that the proposed hybrid approach allows reducing both the MRE and the MAE by 3.7%, from 6.07% to 5.85% and from 663.6 kW to 639.1 kW respectively. The curve fitting is also improved of about 5% (from 44.7% to 46.9%). Although these improvements do not appear to be significant, they validate the last proposed hybrid approach for estimating the thermal power consumption of the hot water distribution network (W_{DN}).

6. Conclusion and perspectives

As part of the OptiEnR project, the present paper, the first of a series of three dealing with the modeling and the optimization of a district boiler, situated at La Rochelle (west coast of France), focuses on forecasting both the outdoor temperature and the thermal power consumption of the hot water distribution network. This district boiler is composed of a wood boiler, a gas-fuel oil boiler, a breaking pressure bottle and a cogeneration plant. With the aim of optimizing the use of a hot water thermal storage unit and reducing the gas-fuel oil boiler coverage rate (i.e. the fossil energy consumption), a model predictive controller will be developed and implemented. It requires forecasting, over the next 4h30, the two above-mentioned parameters.

The proposed MRA-ANN methodology deals with the concept of time series and uses a wavelet-based multi-resolution analysis and artificial neural networks. Multi-layer Perceptrons were trained and validated, using the wavelet decomposition coefficients of past sequences, for estimating both the detail and approximation coefficients of the considered target sequences. Substituting the prediction task of an original time series of high variability by the estimation, using artificial

neural networks, of its wavelet coefficients on different levels of lower variability (according to different frequency domains), while preserving the temporal characteristics of the series, is the main idea of the present work. Then, the reconstruction of the target sequences is performed by simply summing up the estimated coefficients. One can note that the wavelet-based multi-resolution analysis allows isolating both the overall trend and the 24-hour pseudo period characterizing the considered time series (approximation coefficients) from the variability caused by climatic phenomena (detail coefficients). Because these coefficients are of lower variability than the original and to-be-forecasted time series, the MRA-ANN methodology makes the estimation of future values easier.

The number of considered past sequences (M), the sampling time (T), the wavelet decomposition level (N) and the wavelet order (R) impact the forecast accuracy. The networks topology, basically the number of hidden neurons (F), as well as training parameters, such as the number of iterations, the learning rate and the error goal, also impacts the networks generalization capability. That is why the present paper deals with the study and the quantification of all of these influences. It focuses also on the impact of the random initialization of all the networks synaptic weights on the performance of the models and on the suitability of developing specialized models (S). As a result of the study about outdoor temperature forecasting, the optimal configuration, leading to the most accurate forecasts (the MRE, the MAE and the FIT are about 4.2%, 1.2°C and 60% respectively), is defined as follows: $M = 4$, $T = 30$, $N = 5$, $R = 4$, $F = 5$ and $S = 1$. These results validate the proposed MRA-ANN methodology and highlight both the impact of working with wavelet decomposed sequences and the generalization capability of multi-layer neural networks.

Next, two different ways of doing have been considered for forecasting the thermal power consumption of the hot water distribution network: first, using the previously forecasted outdoor temperature (both parameters are strongly correlated) and, secondly, applying the MRA-ANN methodology used to forecast outdoor temperature again (in this case, the optimal configuration highlighted during the first part of the work has been re-used to develop and validate the new model). Finally, a hybrid approach is proposed, combining time series forecasting and linear curve fitting. Whatever the considered approach, the results are satisfactory. The MRE, the MAE and the FIT are ranging between 5.85% and 6.97%, 639.1 kW and 761.9 kW and 37.6% and 46.9% respectively. The hybrid approach provides the best results. However, one can note that, comparing the two forecasted parameters, forecasts accuracy is not as good as that achieved when predicting outdoor temperature. It can be explained by hardly-quantifiable influences leading to amplitude variations. Some peaks of consumption may be due to an unusual functioning of big buildings heating systems (for example, an exhibition hall connected to the hot water distribution network). Some other peaks may be the consequence of the way the thermal power consumption is specified, using both the temperatures of the water entering the distribution network and coming back from this network. So, if the gas-fuel oil boiler starts working, the temperature of the water entering the network will increase quickly but no significant impact will be noticed on the temperature of the cold water coming back to the plant until a few minutes. However, this leads to variations of the thermal power consumption, induced by both the functioning of the district boiler and the thermal inertia of the hot water distribution network. One can suppose that adding to the plant a hot water thermal storage unit (one of the next steps listed in the OptiEnR research project) will limit the presence of hardly-predictable peaks of consumption.

As just mentioned, future work will now focus on optimizing the performance of the district boiler of La Rochelle, adding to the plant a thermal storage unit which will be managed thanks to a model-based predictive controller. Using forecasted outdoor temperature and thermal power consumption sequences, such a controller calculates an optimal command sequence to be applied to the thermal storage feed pump. This allows adjusting the flow of the water passing through this unit. Let us remember that the main objective of the work deals with the minimization of the fossil energy consumption, stocking renewable energy during low-demand periods and using it during high-demand periods. Future papers will describe both the modeling and the way the district boiler functioning will be optimized.

7. References

- [1] Agence Départementale de l'Environnement et de la Maîtrise de l'Énergie, Les chiffres clés du bâtiment en 2006, Publications de l'ADEME, 2007.
- [2] Official Journal of the European Communities, Directive 2002/91/EC of the European Parliament and the Council of 16 December 2002 on the energy performance of buildings, 2002.

- [3] Europe's energy position. Present & future, Market Observatory for Energy, 2008.
- [4] Cofely GDF-SUEZ, <http://www.cofely-gdfsuez.fr/en/homepage/>.
- [5] Weiss France, http://www.weiss-france.fr/index.php?_lngweb=en.
- [6] C.E. Garcia, D.M. Prett, Model predictive control: Theory and practice. A survey, *Automatica* 25 (3) (1989) 335-348.
- [7] S.J. Qin, T.A. Badgwell, A survey of industrial model predictive control technology, *Control Engineering Practice* 11 (2003) 733-764.
- [8] K. Liu, S. Subbarayan, R.R. Shoults, M.T. Manry, C. Kwan, F.L. Lewis, J. Naccarino, Comparison of very short-term load forecasting techniques, *IEEE Transactions on Power Systems* 11 (2) (1996) 877-882.
- [9] A. Khotanzad, M. Davis, A. Abaye, D. Maratukulam, An artificial neural network hourly temperature forecaster with applications in load forecasting, *IEEE Transactions on Power Systems* 11 (2) (1996) 870-876.
- [10] A. Khotanzad, R. Azkhami-Rohani, T. Lu, M. Davis, A. Abaye, D. Maratukulam, ANNSTLF: a neural-network-based electric load forecasting system, *IEEE Transactions on Neural Networks* 8 (4) (1997) 835-846.
- [11] R.C. Hwang, H.C. Huang, Y.J. Chen, J.G. Hsieh, Power load forecasting by neural network with a learning process for considering overtraining problem, *Proceedings of the International Conference on Energy Management and Power Delivery*, 1996, pp. 317-322.
- [12] H. Yoo, R. Pimmel, Short term load forecasting using a self-supervised adaptive neural network, *IEEE Transactions on Power Systems* 14 (2) (1999) 779-784.
- [13] S.S. Sharif, J.H. Taylor, Real-time load forecasting by artificial neural networks, *IEEE Power Engineering Society Summer Meeting*, 2000, pp. 496-501.
- [14] P. González, J.M. Zamarreño, A short-term temperature forecaster based on a novel radial basis functions neural network, *International Journal of Neural Systems* 11 (1) (2001) 71-77.
- [15] P. González, J.M. Zamarreño, A short-term temperature forecaster based on a state space neural network, *Engineering Applications of Artificial Intelligence* 15 (2002) 459-464.
- [16] H.S. Hippert, C.E. Pedreira, R.C. Souza, Combining neural networks and ARIMA models for hourly temperature forecasts, *Proceedings of the IEEE International Joint Conference on Neural Networks*, 2000, pp. 414-419.
- [17] I. Tassadduq, S. Rehman, K. Bubshait, Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia, *Renewable Energy* 25 (2002) 545-554.
- [18] M. Chaabene, M. Ben Ammar, Neuro-fuzzy dynamic model with Kalman filter to forecast irradiance and temperature for solar energy systems, *Renewable Energy* 33 (2008) 1435-1443.
- [19] J.S.R. Jang, ANFIS: Adaptive-Neural-network-based Fuzzy Inference System, *IEEE Transactions on Systems, Man, and Cybernetics* 23 (3) (1993) 665-685.
- [20] J.S.R. Jang, C.T. Sun, Neuro-fuzzy modeling and control, *IEEE Proceedings* 83 (3) (1995) 378-406.
- [21] C.T. Lin, C.S.G. Lee, *Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems*, Upper Saddle River, NJ: Prentice Hall, 1996.
- [22] S. Makridakis, S.C. Wheelwright, R.J. Hyndman, *Forecasting: methods and applications*, Wiley, New York, 1998.
- [23] S.M. Pandit, S.M. Wu, *Time series and system analysis with applications*, New York: John Wiley & Sons, 1983.
- [24] R.E. Kalman, A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* 82 (1) (1960) 35-45.
- [25] D.E. Cathin, *Estimation, control and the discrete Kalman filter*, Berlin: Springer, 1988.
- [26] E.L. Hasltine, J.B. Rawlings, Critical evaluation of extended Kalman filtering and moving-horizon estimation, *Industrial & Engineering Chemistry Research* 44 (8) (2005) 2451-2460.
- [27] S. Grieu, A. Traoré, M. Polit, J. Colprim, Prediction of parameters characterizing the state of a pollution removal biologic process, *Engineering Applications of Artificial Intelligence* 18 (5) (2005) 559-573.
- [28] C.C. Allen, A simplified equation for minimum temperature prediction, *Month. Wea. Rev.* 85 (1957) 119-120.
- [29] A.M. Scheider, T. Takenawa, D.A. Schiffman, 25-hour electric utility load forecasting. In: D.W. Bunn, E.D. Farmer (Eds.), *Comparative Models for Electrical Load Forecasting*, Wiley, New York, 1985, pp. 87-108.

- [30] J.Y. Fan, J.D. McDonald, A real time implementation of short-term load forecasting for distribution power systems, *IEEE Transactions on Power Systems* 9 (2) (1994) 988-994.
- [31] T. Chow, C. Leung, Neural network based short-term load forecasting using weather compensation, *IEEE Transactions on Power Systems* 11 (4) 1736-1742.
- [32] G. Box, G. Jenkins, *Time series analysis: forecasting and control*, San Francisco: Holden-Day, 1970.
- [33] A. Pankratz, *Forecasting with univariate Box-Jenkins models: concepts and cases*, New York: John Wiley & Sons, 1983.
- [34] S.M. Chen, J.R. Hwang, Temperature prediction using fuzzy time series, *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics* 30 (2) (2000) 263-275.
- [35] Q. Song and B.S. Chissom, Fuzzy time series and its models, *Fuzzy Sets and Systems* 54 (3) (1993a) 269-277.
- [36] Q. Song and B.S. Chissom, Forecasting enrollments with fuzzy time series - Part I, *Fuzzy Sets and Systems* 54 (1) (1993b) 1-9.
- [37] Q. Song and B.S. Chissom, Some properties of defuzzification neural networks, *Fuzzy Sets and Systems* 61 (1) (1994a) 83-89.
- [38] Q. Song and B.S. Chissom, Forecasting enrollments with fuzzy time series - Part II, *Fuzzy Sets and Systems* 62 (1) (1994b) 1-8.
- [39] L.W. Lee, L.H. Wang, S.M. Chen, Y.H. Leu, Handling forecasting problems based on two-factors high-order fuzzy time series, *IEEE Transactions on Fuzzy Systems* 14 (3) (2006) 468-477.
- [40] R.E. Abdel-Aal, Hourly temperature forecasting using abductive networks, *Engineering Applications of Artificial Intelligence* 17 (2004) 543-556.
- [41] R. Dong, W. Pedrycz, A granular time series approach to long-term forecasting and trend forecasting, *Physica A: Statistical Mechanics and its Applications* 387 (13) (2008) 3253-3270.
- [42] L.W. Lee, L.H. Wang, S.M. Chen, Temperature prediction and TAIFEX forecasting based on high-order fuzzy logical relationships and genetic simulated annealing techniques, *Expert Systems with Applications* 34 (2008) 328-336.
- [43] S. Kirkpatrick, C.D. Gelatt Jr, M.P. Vecchi, Optimization by simulated annealing, *Science* 220 (4598) (1983) 671-680.
- [44] M. Gen, R. Cheng, *Genetic algorithms and engineering design*, New York: John Wiley & Sons, 1997.
- [45] D.E. Goldberg, *Genetic algorithm in search, optimization, and machine learning*, Massachusetts: Addison-Wesley, 1989.
- [46] D.E. Goldberg, B. Korb, K. Deb, Messy genetic algorithms: motivation, analysis, and first results, *Complex Systems* 3 (5) (1989) 493-530.
- [47] J.H. Holland, *Adaptation in natural and artificial systems*, Massachusetts: MIT Press, 1975.
- [48] E. González-Romera, M.A. Jaramillo-Morán, D. Carmona-Fernández, Forecasting of the electric energy demand trend nad monthly fluctuation with neural networks, *Computers and Industrial Engineering* 52 (3) (2007) 336-343.
- [49] E. González-Romera, M.A. Jaramillo-Morán, D. Carmona-Fernández, Monthly electric energy demand forecasting with neural networks and Fourier series, *Energy Conversion and Management* 49 (2008) 3135-3142.
- [50] S.K. Mitra, J.F. Kaiser, *Handbook for digital signal processing*, New York: John Wiley & Sons, 1993.
- [51] R.E. Abdel-Aal, A.Z. Al-Garni, Forecasting monthly electric energy consumption in Eastern Saudi Arabia using univariate time-series analysis, *Energy* 22 (11) (1997) 1059-1069.
- [52] S. Saab, E. Badr, G. Nasr, Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon, *Energy* 26 (1) (2001) 1-14.
- [53] R. Gao, L.H. Tsoukalas, Neural-wavelet methodology for load forecasting, *Journal of Intelligent and Robotic Systems* 31 (1-3) (2001) 149-157.
- [54] D.M. Tanton, R.R. Cohen, S.D. Probert, Improving the effectiveness of a domestic central-heating boiler by the use of heat storage, *Applied Energy* 27 (1987) 53-82.
- [55] U. Stritih, V. Butala, Optimization of a thermal storage unit combined with a biomass boiler for heating buildings, *Renewable Energy* 29 (2004) 2011-2022.
- [56] M. Cao, J. Cao, Optimal design of thermal-energy stores for boiler plants, *Applied Energy* 83 (2006) 55-68.
- [57] I. Daubechies, *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics, 1992.

- [58] C.E. Heil, D.F. Walnut, Continuous and Discrete Wavelet Transforms, *SIAM Review* 31 (1989) 628-666.
- [59] M.A. Cody, The Fast Wavelet Transform - Beyond Fourier transforms, *Dr Dobb's Journal* (1992) 16-28.
- [60] V. Perrier, T. Philipovitch, C. Basdevant, Wavelet spectra compared to Fourier spectra, *Journal of Mathematical Physics* 36 (1995) 1506-1519.
- [61] J.A. Gubner, W.-B. Chang, Wavelet transforms for discrete-time periodic signals, *Signal Processing* 42 (1995) 167-180.
- [62] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1999.
- [63] D.B. Percival, A.T. Walden, *Wavelet methods for time series analysis*, Cambridge University Press, 2000.
- [64] P.S. Addison, *The illustrated wavelet transform handbook*, Institute of Physics, 2002.
- [65] R. Rosenblatt, *The perceptron: a perceiving and recognizing automaton*, Project PARA, Report 85-460-1, Cornell Aeronautical Laboratory, 1957.
- [66] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys.* 5 (1943) 115-133.
- [67] K. Hornik, M. Stinchcombe, H. White, Multi-layer feedforward networks are universal approximation, *Neural Networks* 2 (1989) 359-366.
- [68] S. Grieu, A. Traoré, M. Polit, J. Colprim, Prediction of parameters characterizing the state of a pollution removal biologic process, *Engineering Applications of Artificial Intelligence* 18 (2005) 559-573.
- [69] C. Charalambous, Conjugate gradient algorithm for efficient training of artificial neural networks, *IEEE Proc.* 139 (1992) 301-310.
- [70] M.T. Hagan, M. Menhaj, Training feedforward networks with the Marquardt algorithm, *IEEE Trans. Neural Networks* 5 (1994) 989-993.
- [71] P.J. Brockwell, R.A. Davis, *Time series: theory and methods*, 2nd edition, Springer, 1991.
- [72] P.J. Brockwell, R.A. Davis, *Introduction to time series and forecasting*, Springer, 1997.
- [73] V.G. Tran, S. Grieu, A. Traoré, F. Thiéry, M. Estaben, M. Polit, Wavelets decomposition and Box-Jenkins methodology or neural networks for short-term electric consumptions forecast, *International Conference on Renewable Energy and Eco-Design in Electrical Engineering iREED 2008*, Montpellier, France, 2008.
- [74] T. Nengling, J. Stenzel, W. Hongxiao, Techniques of applying wavelet transform into combined model for short-term load forecasting, *Electric Power Systems Research* 76 (2006) 525-533.
- [75] S. Grieu, V.G. Tran, M. Polit, Q.T. Tran, Transformée en ondelettes discrète et réseaux de neurones artificiels pour la prédiction d'irradiation solaire à court terme, *Energies renouvelables et éco-conception en génie électrique, Revue de l'Electricité et de l'Electronique* 5 (2009) 52-60.
- [76] V.G. Tran, S. Grieu, M. Polit, Q.T. Tran, Forecasting of solar irradiation using wavelets decomposition and cascade-correlation neural networks, *International Conference on Renewable Energy and Eco-Design in Electrical Engineering iREED 2008*, Montpellier, France, 2008.
- [77] V.G. Tran, S. Grieu, M. Polit, Q.T. Tran, H.L. Tran, Forecasting of wind speed using wavelets analysis and cascade-correlation networks, *European Wind Energy Conference EWEC 2009*, Marseille, France, 2009.