

# Forecasting Urban Water Demand Via Wavelet-Denoising and Neural Network Models. Case Study: City of Syracuse, Italy

Salvatore Campisi-Pinto · Jan Adamowski ·  
Gideon Oron

Received: 7 December 2010 / Accepted: 5 June 2012 /  
Published online: 4 July 2012  
© Springer Science+Business Media B.V. 2012

**Abstract** Forecasting urban water demand can be of use in the management of water utilities. For example, activities such as water-budgeting, operation and maintenance of pumps, wells, reservoirs, and mains require quantitative estimations of water resources at specified future dates. In this study, we tackle the problem of forecasting urban water demand by means of back-propagation artificial neural networks (ANNs) coupled with wavelet-denoising. In addition, non-coupled ANN and Linear Multiple Regression were used as comparison models. We considered the case of the municipality of Syracuse, Italy; for this purpose, we used a 7 year-long time series of water demand without additional predictors. Six forecasting horizons were considered, from 1 to 6 months ahead. The main objective was to implement a forecasting model that may be readily used for municipal water budgeting. An additional objective was to explore the impact of wavelet-denoising on ANN generalization. For this purpose, we measured the impact of five different wavelet filter-banks (namely, *Haar* and

---

S. Campisi-Pinto  
Albert Katz International School for Desert Studies, The Jacob Blaustein Institutes for Desert Research,  
Ben-Gurion University of the Negev, Beersheba, Israel  
e-mail: campisi@post.bgu.ac.il

J. Adamowski (✉)  
Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill  
University, Montreal, Canada  
e-mail: jan.adamowski@mcgill.ca

G. Oron  
The Department of Industrial Engineering and Management, Ben-Gurion University of the Negev,  
Beer Sheva 84105, Israel  
e-mail: gidi@post.bgu.ac.il

Daubechies of type *db2*, *db3*, *db4*, and *db5*) on a single neural network. Empirical results show that neural networks coupled with *Haar* and Daubechies' filter-banks of type *db2* and *db3* outperformed all of the following: non-coupled ANN, Multiple Linear Regression and ANN models coupled with Daubechies filters of type *db4* and *db5*. The results of this study suggest that reduced variance in the training-set (by means of denoising) may improve forecasting accuracy; on the other hand, an oversimplification of the input-matrix may deteriorate forecasting accuracy and induce network instability.

**Keywords** Artificial neural networks · Wavelets · Denoising · Forecasting · Water demand

## 1 Introduction

Short-term urban water forecasting can be of use for the efficient operation and maintenance of pumps, wells, reservoirs and mains, whereas medium and long term forecasting may be required for infrastructure investments and regional policy making (Herrera et al. 2010). In this context, modeling and scenario-building becomes critical, especially when conditions experienced in the past are not expected to repeat themselves in the future. This may be the case with fast demographic growth, climate changes and over-pumping from aquifers, which are likely to impact not only the amount but also the quality and the patterns of future water supply and demand (Issar and Zohar 2009; Oron et al. 1999).

According to Jain and Ormsbee (2002), perhaps the most frequently used forecasting models for water demand are based on linear regression and time series analysis. In the 1980s, Maidment and Miaou (1986) used Box and Jenkins models to forecast daily municipal water use as a function of rainfall and air temperature in nine US cities, and Smith (1988) included day-of-week effects among the predictors and randomly varying means as a regression method. An et al. (1996) revisited rough set methodologies describing the relationships between water factors and water consumption. On the other hand, in the field of deterministic prediction, important research has been conducted in developing models of water budgeting according to different foundations such as classical hydrology (Grubbs 1994) and environmental isotopes (Adar et al. 1988).

Since the introduction of the back propagation algorithm by Rumelhart et al. (1986), Artificial Neural Networks (ANNs) have been used in a large variety of forecasting applications. This method has been frequently used because it can tackle problems where the underlying relationships in the time series are complex or unknown, but there is enough data (i.e. observations) to train a network (Zhang et al. 1998). A number of ANN configurations have been used for hydrological forecasting with good results. In this respect, Jain and Ormsbee (2002), Bougadis et al. (2005), Adamowski (2008), Adamowski and Karapataki (2010), Güldal and Tongal (2010) and Shirsath and Singh (2010) used ANNs to forecast various aspects of water resources demand. In all these instances, it was concluded that ANN-based methods provide better results than conventional autoregressive methods. Zhang et al. (1998), in a state-of-the-art survey, observed that although some contradicting reports exist in the literature, ANN is a useful method for non-linear modeling while autoregressive approaches may be more suitable for linear relationships in the data.

When comparing ANNs to newer methods mixed results were found. Msiza et al. (2007) compared ANNs to support vector machines (SVM) for urban water demand, concluding that ANNs were more precise; on the other hand, Herrera et al. (2010)

used ANN, SVM, projection pursuit regression and multivariate adaptive regression splines, and found that SVMs were more precise for urban water demand forecasting.

With regards to spectral analysis of time series, the discovery of wavelet functions has shed new light onto the analysis of non-stationary and/or noisy phenomena (Rioul and Vetterli 1991). In geophysics it is not uncommon to find time series that include a trend in the mean, non-constant variance, and discontinuities as well as background noise (Jowitt and Xu 1992). Foufoula-Georgiou and Kumar (1994) described the basic properties that make wavelet analysis a powerful tool for geophysical applications. It was noted that the multi-resolution approach may be useful for the analysis of multi-scale features, detection of singularities, analysis of transient phenomena, non-stationary series, and fractal processes.

With specific attention to hydrology, Labat (2005) noted that the application of wavelets may lead to several improvements in the analysis of global hydrological fluctuations and their mutual time varying relationships. It was also suggested that wavelets should be used more systematically, notably in hydrology as a preferable alternative to classical Fourier analysis. For example, Chou (2011) used wavelet-denoising instead of Fourier thresholding in order to eliminate the effect of high-tide low-tide variations at gauge level from the relation between rainfall and run-off. Adamowski et al. (2009) developed a wavelet-aided technique for trend detection in monthly streamflow and water resource forecasting and Labat (2008) used wavelets for multi-scale analysis of the world's largest river discharges.

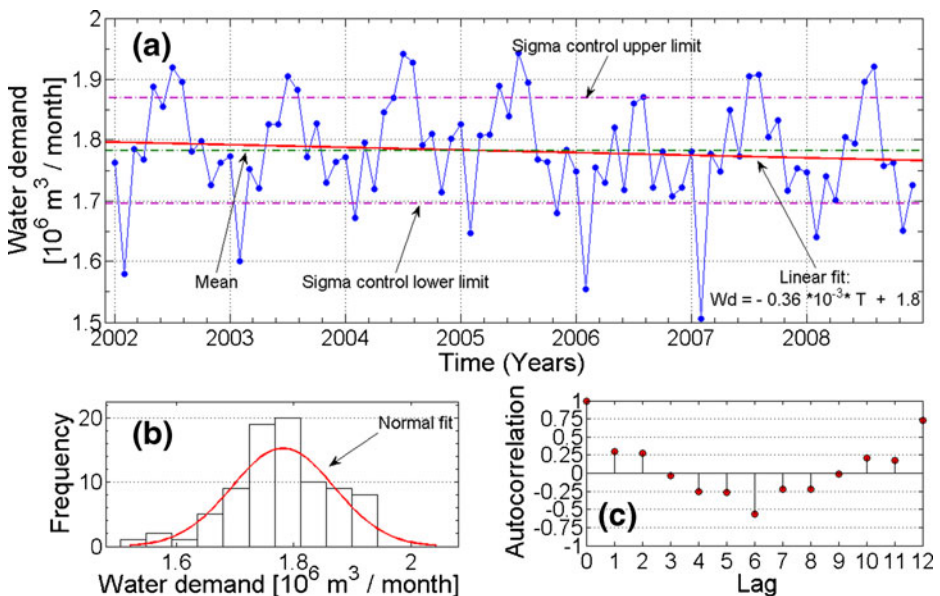
In recent years, several authors noted that coupling wavelet transforms with ANNs could provide highly accurate hydrological forecasts. Partal and Cigizoglu (2008) used wavelets in conjunction with ANNs for the forecasting of evapotranspiration; Zhou et al. (2007) and Kisi (2008, 2009) used wavelet-decomposition in order to improve ANN forecasting of river stream-flow; Wang et al. (2009) analyzed time series by means of wavelet-transforms in order to forecast “xunly” mean discharge of the Three Gorges Dam (where “xun” is the Chinese term indicating a 10-day period); Adamowski and Sun (2010) used coupled wavelet-neural network models in flow forecasting; Adamowski and Chan (2011) used wavelets and ANN for the forecasting of ground-water levels; and Adamowski et al. (2012) used wavelets and neural networks to forecast urban water demand for the city of Montreal, Canada. These studies found that coupled wavelet-neural network models generally provided more accurate forecasts than other models (such as ARIMA, ANN, MLR, etc.). With specific regard to denoising methods based on wavelets, Nourani et al. (2009) and Cannas et al. (2006) both explored the multi-scaling property of wavelets for maximization of ANN forecasting accuracy (in the context of flow forecasting). Empirical results show that networks trained with pre-processed data performed better than networks trained on un-decomposed, noisy raw signals.

In this study, we tackle the problem of forecasting urban water demand by means of back-propagation artificial neural networks coupled with wavelet-denoising. In addition, non-coupled ANN and Linear Multiple Regression methods were used as comparison models. We considered the case of the municipality of Syracuse, Italy. For this purpose, we used a 7 year-long time series of monthly urban water demand; no additional predictors such as demographic, economic or environmental variables were used. ANN tuning is fully described in a step by step procedure. Six forecasting horizons, from 1 to 6 months ahead, were considered. The main objective of the study was to calibrate a forecasting model for operational applications (i.e. monthly water budgeting for the municipality of Syracuse, Italy). Moreover, we measured the impact of different filter-banks (namely, *Haar* and Daubechies of type *db2*, *db3*, *db4*, and *db5*) on a single back-propagation multi layered neural network for the prediction of monthly water

demand. Thus, an additional objective was to explore the impact of different wavelets on ANN generalization.

## 2 Data

The dataset is a time series consisting of 84 measurements (from January 2002 to December 2008) of monthly water consumption of the municipality of Syracuse, Italy (Fig. 1a). It cumulates monthly water flow measured at the outlet of several urban water tanks. Syracuse is located in the South-East coast of Sicily at 37°5'0" North and 15°17'0" East, according to the World Geodetic System 84. The Mediterranean climate is characterized by mild, wet winters and warm to hot, dry summers (Koppen climate classification: Csa). The time series in Fig. 1a shows seasonal oscillation around the mean; it reaches a local maximum approximately every 12 months during the summer and a local minimum during the winter. The mean of the year-maximum values is 1.92 million m<sup>3</sup>/month, the mean of the year-minimum is 1.60 million m<sup>3</sup>/month. The mean monthly water consumption is 1.78 million m<sup>3</sup>/month with a standard deviation of 0.086 million m<sup>3</sup>/month. Mean and standard deviation are not constant; a decreasing trend is also detectable. The samples come from an underlying sinusoidal model and they have a high degree of autocorrelation at 6 and 12 lagged observations (Fig. 1c). The Lilliefors test (Conover 1980) rejects the null hypothesis at the 5 % significance level that the sample in our time series comes from a distribution in the normal family (against the alternative that it does not come from a normal distribution). Thus, the distribution is not likely to be normally distributed (Fig. 1b).



**Fig. 1** Monthly water demand of the municipality of Syracuse, Italy. In (a) 84 measurements in blue; the central dotted line (in green) represents the overall mean; the external dotted lines (in violet) represent the control limits, one standard deviation above and below the mean; the continuous line (red) represents the trend line. In (b) the histogram of water demand versus the normal fit; in (c) the autocorrelation coefficients are included

### 3 Methodology

#### 3.1 Multiple Linear Regression

Multiple linear regression (MLR) models are capable of making predictions based on multiple inputs. The analytical expression (Holder 1985) is defined by:

$$y_t = a_0 + a_1x_1(t) + \dots + a_nx_n(t) + \varepsilon(t) \quad (1)$$

where  $y_t$  is the output,  $a_0$  is a constant,  $a_1 \dots a_n$  are regression coefficients computed via the least squares method,  $x_1(t) \dots x_n(t)$  are the inputs into the model, and  $\varepsilon(t)$  is a random variable with zero mean and constant variance.

#### 3.2 Artificial Neural Networks

An artificial neural network (ANN) is an interconnected group of simple processing units mimicking the function of biological neurons first discovered in physiology. According to the prevalent terminology, we call neurons or nodes the processing units of an artificial neural network. Neurons are arranged in layers; neurons between layers are connected by links called weights. Error-back propagation includes an additional set of feed-back connections from the output to the input layer (Rumelhart et al. 1986). Back propagation networks are a generalization of the Widrow-Hoff learning rule (Widrow and Lehr 1990), in which the network weights are adapted along the negative gradient of the performance function (Rumelhart et al. 1986). Back-propagation multi-layered fully connected neural networks are among the most popular and proven (Hagan et al. 1996). Complete derivation of the model and the learning algorithm can be found in Rumelhart et al. (1986) and Haykin (1994). A single-output multilayer feed forward back propagation network performs the following mapping from the input to the output data (Zhang et al. 2001):

$$\hat{y}_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-n}) \quad (2)$$

where  $\hat{y}_t$  is the output at time  $t$ ,  $n$  is the dimension of the input vector or the number of past observations used to predict future observations, and  $f$  is a nonlinear function determined both by the network and the data at hand. From Eq. 2 the feed-forward function can be viewed as a general nonlinear, autoregressive model (Zhang et al. 2001). In a trained network, the relation between outputs and inputs has the following expression (Zhang 2003):

$$y_t = \alpha_0 + \sum_{j=1}^q \alpha_j g \left( \beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-i} \right) + \varepsilon_t \quad (3)$$

where  $\alpha_j$  for  $(j=0,1,2,\dots,q)$  and  $\beta_{ij}$  for  $(i=0,1,2,\dots,p; j=0,1,2,\dots,q)$  are the model parameters, also referred to as weights in the connectionist literature;  $p$  is the number of input nodes and  $q$  is the number of hidden nodes; the function  $g$  is a logistic function, in our case we used a log-sigmoid function:

$$g(x) = \frac{1}{1 + e^{(-x)}} \quad (4)$$

Neural networks operate on the basis of learning rules which define exactly how the network weights should be adjusted (updated) between successive training cycles (epochs). Thus, training is the process of featuring the function  $f$  in Eq. 2 which eventually is uniquely

determined by the linking weights of the network. We used a supervised learning method. In practice the network learned the mapping from the input data space to the output, by means of a set of correct solutions used for supervising the training process. We used a Levenberg-Marquardt (LM) training algorithm. LM is often the fastest back propagation algorithm, and is highly recommended as a first-choice for supervised learning (Hagan et al. 1996). The mean squared error (MSE) was used as a performance function. We used a log-sigmoid activation function (Eq. 4) for all the hidden layers, and a linear activation for the output layer. Training was terminated when any of the following conditions occurred: a) the maximum number of epochs (repetitions) was reached, in this case  $\text{maximum}(\text{epoch}) = 100$ ; b) the maximum amount of time had been exceeded, in this case  $\text{maximum}(\text{time}) = \text{infinite}$ ; c) performance had been minimized to the goal, in this case no goal was pre-determined,  $\text{goal} = \text{null}$ ; d) the performance gradient fell below a threshold,  $\text{minimum}(\text{gradient}) = 1e^{-10}$ ; e) “ $\mu$ ” exceeded  $\text{maximum}(\mu) = 1e^{10}$ ; f) validation performance increased more than 5 times since the last time it decreased (over the validation dataset).

At each ANN iteration, the training set was randomly partitioned into three subsets: training, validation, and testing. The training-subset was used to update the weights of the network. The validation-subset was used to set a rule to interrupt training (as described above). The test-subsets were used to measure ANN performances on a portion of data that had not been used for training, also referred to as off-set data. There are several heuristics for partitioning: we used 60 % of the dataset for training, 20 % for validation and 20 % for testing (Haykin 1994). In practice, at each ANN iteration, we forecasted only 14 months out of 72 dates comprising the forecasting domain (from January 2003 to December 2008). By means of several hundred repetitions we managed to cover the entire forecasting domain with at least 30 off-set forecasts for each single month.

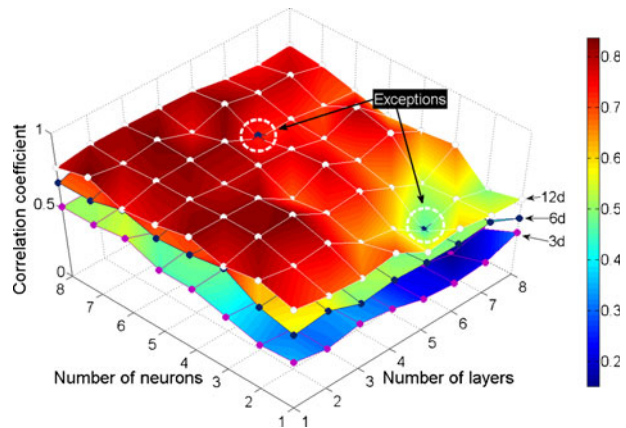
The next steps of ANN configuration consisted of choosing: a) the dimension of the input vectors; b) the number of layers; c) the number of neurons per layer. Regarding the optimal length of the input vectors, the objective was to determine the optimal number of lagged observations, which is  $n$  in Eq. 2. This is among the tasks with the largest impact on convergence (Haykin 1994). There is no theoretical support that can be used to guide the selection of  $n$  (Zhang et al. 2001). We tested 3, 6 and 12-dimensional input-vectors on a number of networks having hidden layers ranging from 1 to 8 and neurons per layer ranging from 1 to 8, for a total of  $8 \times 8 = 64$  network configurations. We evaluated the performance of each configuration over 100 numerical experiments, for a total of  $3 \times 64 \times 100 = 19,200$  numerical experiments. Figure 2 shows three correlation surfaces corresponding to our 64 configurations trained respectively by 3, 6, and 12-dimensional input vectors. In all cases, other than two exceptions, a 12-dimensional input vector showed higher correlations than any other ANN trained with 6d and 3d input vectors. Hence, we selected a 12d input vector ( $n=12$ ).

In the remaining sections we will refer to a number of variables defined as follows:  $(Y_t)$  is the forecasted value for the month  $t$ ;  $[Y_r]$  is the output vector consisting of the scalars  $(Y_t)$ , where  $r=1, 2, \dots, z$  is the ordinal number of the experiment; and  $[Y_m]$  is the vector of final forecasts of the model  $m$ , defined as the average of the set  $\{Y_r\}$ . The model performances are measured by a number of indexes based on the benchmarking of  $[Y_m]$  against targets  $[T]$ , where  $[T]$  is the vector of measured monthly water demand.

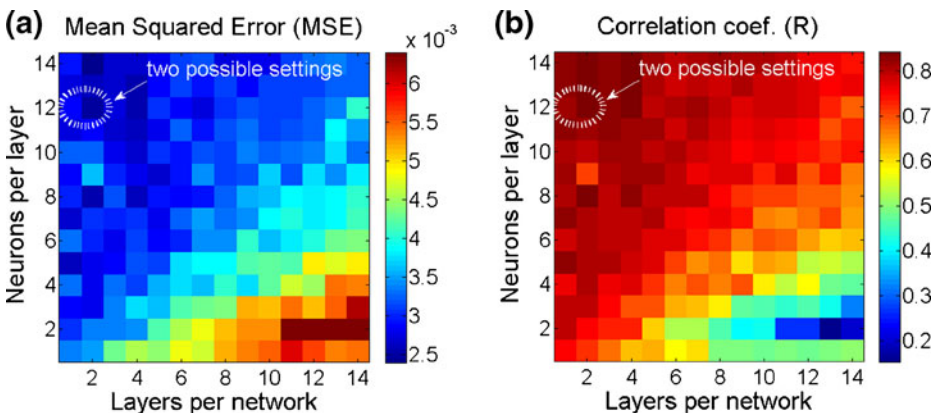
Regarding the optimal number of layers and neurons, in most function approximation tasks, one layer (with several neurons) is sufficient to approximate continuous functions; generally, two hidden layers may be necessary for learning functions with discontinuities (Hagan et al. 1996). Nevertheless, these general considerations may not necessarily be optimal for the data set at hand. Several approaches are suggested to define neural architecture (Widrow and Lehr 1990). Guidelines may be either heuristic or derived from empirical



**Fig. 2** Correlation surface of different ANNs. On the horizontal grid the combination of neurons and layers are presented; on the vertical axis the correlation coefficient corresponding to each combination is presented. Each stratum of the figure corresponds to 3, 6, and 12 dimensional input vectors, respectively



experiments. The methodological “freedom” in the choice of layers and neurons has been mentioned by Zhang et al. (1998) as one of the reasons for inconsistent reports about ANN performances in the scientific literature. Having fixed the length of the input vector equal to 12, we systematically evaluated all the configurations comprising a number of layers ranging from 1 to 14 consisting respectively of a number of neurons ranging from 1 to 14 per layer, for a total of  $14 \times 14 = 196$  configurations. We tested each configuration 100 times, for a total of  $196 \times 100 = 19,600$  numerical experiments. Results are presented in a matrix pictorial form (Fig. 3). Each pixel corresponds to a combination of layers and neurons, the color-bar in Fig. 3a represents the mean squared error (MSE) between outputs  $[Y_m]$  and targets  $[T]$ , and the color bar in Fig. 3b represents the correlation coefficients (R) between outputs  $[Y_m]$  and targets  $[T]$ . It was found that for the dataset at hand, the network configuration that minimized the MSE and maximized the correlation coefficient (R) was the one having 2 layers. More precisely: 12 input nodes in the input layer ( $q=12$  in Eq. 3) and 12 hidden nodes in a single hidden layer ( $p=12$  in Eq. 3), for a total of 24 nodes distributed in 2 layers.



**Fig. 3** Performances of 196 different ANNs. Each pixel represents a different ANN-setting resulting from the combinations of a varying number of layers-per-network (ranging from 1 to 14) and neurons-per-layer (ranging from 1 to 14). The color bars represent the magnitude of the Mean Square Errors (a) and the Correlation Coefficients (b). In all instances ANNs were trained with 12 dimensional input vectors

**Table 1** Network stability as a function of number of layers per network

	Number of layers per network													
	1 <sup>a</sup>	2	3	4	5	6	7	8	9	10	11	12	13	14
Standard deviation of R	0.03	0.06	0.04	0.06	0.07	0.09	0.09	0.11	0.13	0.13	0.16	0.15	0.17	0.17
Standard deviation of MSE	0.24	0.78	0.44	0.55	0.55	0.65	0.67	0.78	0.88	0.88	1.13	1.02	1.05	1.06

<sup>a</sup> optimal combination

An additional test was performed in order to establish how stable this configuration was. For this purpose, we measured the magnitude of the standard deviations of the outputs as a function of number of layers; the larger the standard deviation the less stable the network. Experimental results (Table 1) show that output standard deviation increased with the number of layers. Thus, although in the two layer configuration one can find the best forecasting accuracy, the one layer configuration was more stable without a significant decrease in network performance. Therefore, a one-layer network was chosen to be the optimal configuration.

Finally, within a one layer configuration, we focused on the optimal number of neurons. It was found that 12 neurons consistently returned above average results both in terms of correlation coefficient and mean squared error of outputs versus target. Therefore, this was adopted as the definitive configuration for this study.

To summarize, by means of a step-wise screening where at each step a single parameter was analyzed, we concluded that the network to be used in the remaining part of the study was the one having the following major features: feed-forward back-propagation network with 12 neurons in one hidden layer, fully connected by tan-sigmoid threshold functions to an output layer provided by a linear threshold, trained with a Levenberg-Marquardt algorithm.

### 3.3 Wavelet-Denoising

Wavelets are orthogonal bases of finite length that can be used to represent a time series into a time-scale domain at different resolutions. In this study, wavelet transforms were used for their denoising capabilities in order to improve ANN forecasting accuracy. Wavelets are described by Compo and Torrence (1998) as a “tool to analyze variation of power within a time series”. Because of the compact support in which wavelets are defined, wavelet filter-banks are also well suited to decompose, manipulate and represent non-stationary time series. A practical guide to wavelet-analysis is provided by Compo and Torrence (1998), fundamental manuals are provided by Blatter (1998), Daubechies (1992), Meyer (1992), Strang and Nguyen (1996), Holschneider (1995), and Mallat (1999).

Wavelet analysis consists of decomposition, and wavelet synthesis consists of reconstruction of a given signal. In the continuous domain, the analysis starts from choosing a mother wavelet ( $\psi$ ). The continuous wavelet transform (CWT) is then defined by the integration over all time of the signal multiplied by scaled and shifted versions of the mother wavelet (Mallat 1999):

$$CWT_x^\psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t) \psi^* \left( \frac{t - \tau}{s} \right) dt \quad (5)$$



where  $s$  is the scale parameter,  $\tau$  is the translation and  $*$  stands for the complex conjugate. The CWT produces a continuum of all scales as the output. In the discrete domain, let  $s = s_0^j$  and  $\tau = \tau_0^k$ . The discrete wavelet given by (Mallat 1999) is then defined as follows:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{|s_0^j|}} \psi^* \left( \frac{t - k\tau_0^j s_0^j}{s_0^j} \right) \quad (6)$$

where  $j$  and  $k$  are integers and  $s_0 > 1$  is a fix dilation step. The most practical choice for the parameters  $s_0$  is 2 and  $\tau_0$  is 1. It results in a power of two logarithmic scaling, also referred to as a dyadic grid arrangement of the mother wavelet. The dyadic wavelet can be written in more compact notation as:

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \quad (7)$$

such that  $\{\psi_{j,k}\}_{(j,k) \in \mathbb{Z}^2}$  forms an orthonormal basis for  $L^2(\mathcal{R})$ , the vector space of measurable, square integrable one-dimensional functions. One of the inherent challenges of using the discrete wavelet transform (DWT) for forecasting applications is that it is not shift invariant (i.e. if we shift the beginning of our time series, all of the wavelet coefficients will change). To overcome this problem, a redundant algorithm, known as the “à trous” algorithm was used in this study. Finally, the wavelet can be used as a low-pass filter to prevent any future information from being used during the decomposition. For more information on the issue of shift invariance and causal-filtering, readers are directed to Renaud et al. (2005).

The transform can be presented according to Mallat (1999):

$$c_{i+1}(k) = \sum_{-\infty}^{+\infty} h(t)x_i(k + 2^i t) \quad (8)$$

where  $h$  are the coefficients of the filters of choice,  $x_i(t)$  is the original time series and  $c_{i+1}$  are the coefficients of the wavelet analysis. In practice, in the discrete domain, we implemented a two-channel sub-band coding using quadrature mirror filters (QMFs). This consisted in the analysis of the signal by different filter banks (Low-pass and High-pass filters). In the two band decomposition scheme, starting from  $x(t)$ , the first step produces two sets of coefficients: approximation coefficients (cA1), and detail coefficients (cD1). These vectors are obtained by convolving  $x(t)$  with the low-pass filter Lo\_D, and with the high-pass filter Hi\_D, followed by dyadic decimation. A dyadic decimation must be applied at each step (i.e. downsampling). In practice the decomposition can continue iteratively on the low-pass transform cA1. Several levels of wavelet decomposition were tested; the first level was chosen as the most adequate for our denoising purposes. The inverse transform (IDWT) started from cA1, thus the signal was reconstructed backwards and up-sampled. The algorithm guaranties perfect reconstruction unless the coefficients are somehow manipulated. In our case we suppressed the coefficients of the high-pass transform (i.e. the details at level 1). In this study we used a number of compactly supported discrete wavelet-functions of the type: *Haar*, *db2*, *db3*, *db4*, and *db5* named respectively after Alfred Haar and Ingrid Daubechies (Daubechies 1992). The *Haar* wavelet family is the simplest; it redefines the original signal in terms of averages and differences:

$$\psi(x) = 1 \quad \text{if } x \in [0, 0.5) \quad (9)$$

$$\psi(x) = -1 \quad \text{if } x \in [0.5, 1) \quad (10)$$

$$\psi(x) = 0 \quad \text{if } x \notin [0, 1) \quad (11)$$

$$\phi(x) = 1 \quad \text{if } x \in [0, 1) \quad (12)$$

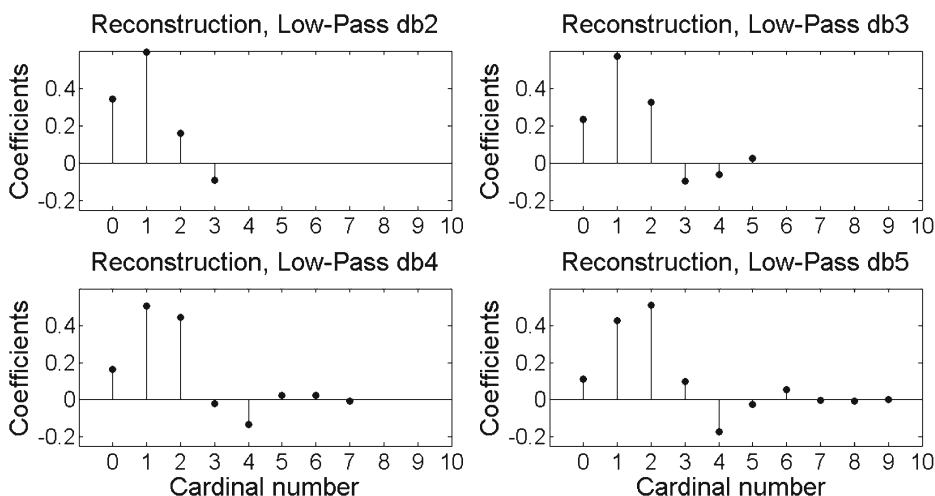
$$\phi(x) = 0 \quad \text{if } x \notin [0, 1] \quad (13)$$

Daubechies wavelets (*dbN*) have no explicit expression except for *db1*, which actually is the above mentioned *Haar* wavelet. They are ranked in terms of their vanishing moments which are equal to  $2N-1$ ; we used  $N$  ranging from 2 to 5, thus wavelet types *db2*, *db3*, *db4*, *db5* (Fig. 4). When non-normalized, *dbN* are Finite Impulse Response (FIR), low pass-filters, of length  $2N$ , of sum 1, of norm  $1/\sqrt{2}$ .

From *Haar* and *dbN* we defined four FIR filters, of length  $2N$  and of norm 1. Once five filter-banks were constructed the decomposition started from the original signal (84 data points of monthly water supply). Usually, the DWT is defined for sequences with length of some power of 2. In order to get a perfect reconstruction we used a symmetrization method for extending our signal on the boundaries to the next power of 2 of the time-series length. Symmetrization assumes that signals can be recovered outside their original support by symmetric boundary value replication. Details on the rationale of these schemes are given in Strang and Nguyen (1996).

### 3.4 One Step and Multi-Step Ahead Forecasting

Three groups of models were constructed:  $\text{WANN}_{\text{dbN}}$ ,  $\text{ANN}_{\text{Raw}}$  and MLR. The  $\text{WANN}_{\text{dbN}}$  implies the alternative use of five different filter-banks: *Haar*, *db2*, *db3*, *db4*, *db5*. The



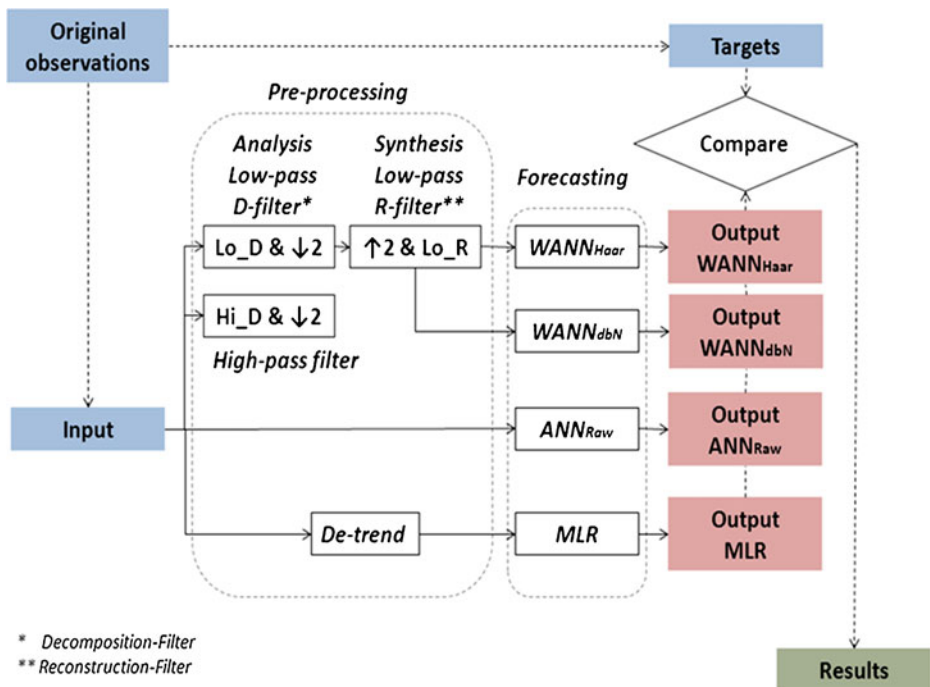
**Fig. 4** Wavelet coefficients for low-pass reconstruction

ANN<sub>Raw</sub> is the non-denoised Artificial Neural Network that is used as a comparison model, as is the Multiple Linear Regression (MLR) model. Having considered that ANNs are embedded with a slight randomness in the initial attribution of weights that ultimately affect predictions, each model was reiterated 200 times with the aim of approaching statistically significant results.

Model simulations begin from the original time series which is used both for building input matrices, as well as target vectors for model benchmarking (Fig. 5). In the WANN models, input matrices were filtered by Low-pass and High-pass filter-banks. High-pass decompositions were suppressed; Low-pass reconstructions were fed into the WANN models. The outputs of 200 reiterations per model were averaged, thereafter benchmarked against targets (i.e. “Compare” in Fig. 5) and ultimately saved as model results. It must be noted that given that only off-set output were used to infer model performances, and considering that off-set outputs can cover only 20 % of the forecasting domain at each model reiteration, the combined result of 200 reiterations per model provided about 30 different off-set forecasting per month for all the months of the forecasting domain.

#### 4 Model Comparison

We used several indices to obtain our estimates of model performances. Indices were selected to meet two objectives: (i) measuring the overall performance of the model over several reiterations; and (ii) comparing results among different models. The



**Fig. 5** Flow chart including data source, data pre-processing, forecasting, model comparison against targets and final results

following indices were used: Fractional Standard Error (FSE), the coefficient of determination ( $R^2$ ), the correlation coefficient (R), the Nash-Sutcliffe model efficiency (E), and indication of the overall model performance was assessed using bias and the bias indicator (B). In particular:

(i) Fractional Standard Error (FSE) is the RMSE divided by the corresponding mean of the targets (observed values). It is a scalable measure of model precision, expressed as:

$$FSE = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n (T_i - Y_i)^2}{\bar{T}}} \quad (14)$$

where  $T_i$ ,  $Y_i$ , and  $\bar{T}$  are the observed, forecasted, and mean of the observed monthly water consumption, respectively. The model becomes more precise as the FSE reaches zero.

(ii) The correlation coefficient (R), expressed as:

$$R = \frac{\sum_{i=1}^n (T_i - \bar{T})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (T_i - \bar{T})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (15)$$

The coefficient R shows how much variability in the data set is accounted for by the model and provides a measure of how likely future outcomes will be forecasted. Values for R range from  $-1$  to  $1$ , with  $1$  referring to maximal correlation.

(iii) The Nash-Sutcliffe model efficiency (E), expressed as:

$$E = 1 - \frac{\sum_{i=1}^n (T_i - Y_i)^2}{\sum_{i=1}^n (T_i - \bar{T})^2} \quad (16)$$

E is used widely in hydrology because it measures the ability of the model to forecast values different from the mean. Values of E range from  $-\infty$  to  $1$ , with  $1$  showing perfect model performance.

(iv) Bias (B), expressed as:

$$B = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n T_i} \quad (17)$$

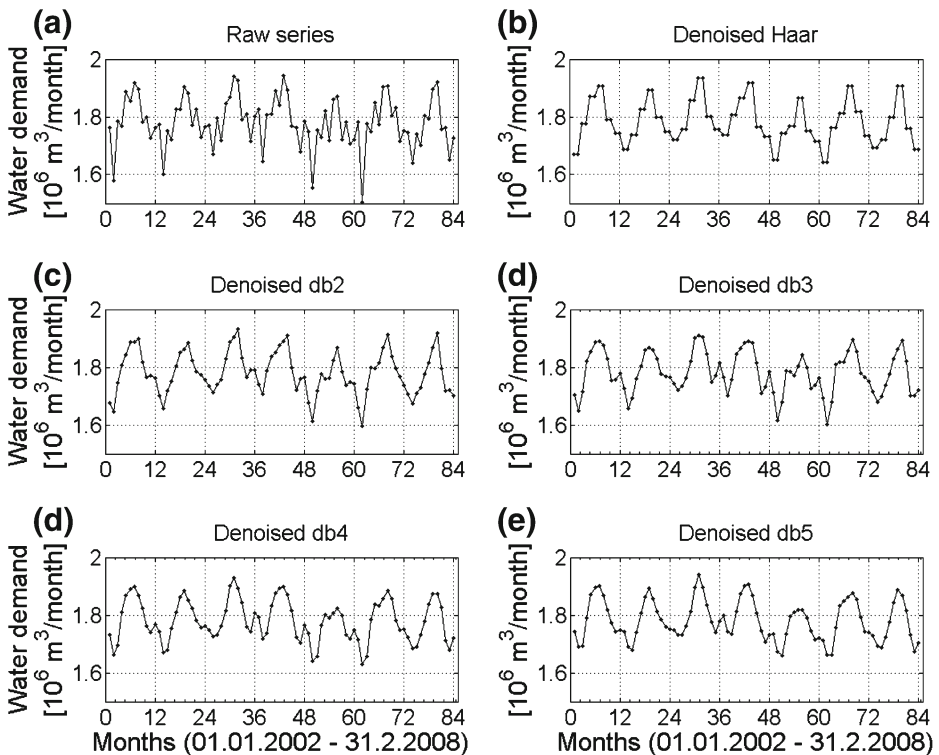
The index B provides a good measure of whether the model is overestimating ( $B > 1$ ) or underestimating ( $B < 1$ ) compared to observed values.  $B = 1$  indicates non-biased model performance.

## 5 Results

### 5.1 Denoising

The original time-series of monthly water demand was denoised by means of six wavelet filter-banks, namely, *Haar*, *db2*, *db3*, *db4*, and *db5*. The denoised versions are reconstructions of the original signal after the high-pass transforms were suppressed at the first level of decomposition (Fig. 6). Several levels of decomposition were tested but the first level was sufficient for suppressing the highest frequencies of the parent time-series. Higher levels of decomposition would be suitable for multi scale analysis. In this case wavelets were used as a denoising tool only; representing the parent series into different scales was not one of the objectives of the study.

Different filter-banks had different impacts on the original parent signal (Table 2). It was noted that: i) the parent mean is preserved by all filter-banks; ii) the variance of the reconstructions decreased when the number of vanishing moments in the filter-banks increased; iii) in all instances the data range of the reconstructed signals was smaller than the data range of the parent signal; iv) the norm of the difference between the parent and denoised versions increased with the number of vanishing moments in the filter-bank; and; v) the correlation coefficients (*R*) decreased when the number of vanishing moments in the filter-banks increased.



**Fig. 6** Signal denoising. In (a) the original signal is presented, in the remaining plots the denoised time-series according to different filter-banks are presented

**Table 2** Statistics of parent and denoised time-series

Filter-bank	Reconstructed time-series					Parent time-series
	Haar	db2	db3	db4	db5	None
$R^a$	0.87	0.84	0.83	0.83	0.81	1
Norm of difference <sup>a</sup>	0.388	0.414	0.432	0.437	0.456	0
Mean	1.78	1.78	1.78	1.78	1.78	1.78
Maximum	1.93	1.93	1.91	1.93	1.94	1.94
Minimum	1.64	1.59	1.60	1.62	1.65	1.50
Range	0.29	0.33	0.31	0.30	0.29	0.43
Variance	0.0057	0.0055	0.0053	0.0052	0.0050	0.0075

<sup>a</sup>reconstructed times-series vs. original time-series

## 5.2 Forecasting

We first present results for 1 month-ahead forecasting, thereafter multi-months ahead forecasting. The non-coupled ANN<sub>Raw</sub> and MLR models had correlation coefficients  $R=0.84$  and  $0.77$  respectively, and fractional standard error  $FSE = 2.6\%$  and  $3.3\%$ , respectively. On the other hand the best coupled model, WANN<sub>Haar</sub>, had correlation coefficient  $R=0.91$  and  $FSE = 2\%$ . Models WANN<sub>db2</sub> and WANN<sub>db3</sub>, although slightly less accurate than the WANN<sub>Haar</sub>, outperformed ANN<sub>Raw</sub>. ANNs coupled with *db4* and *db5* filter-banks did not outperform the non-coupled ANN (Table 3).

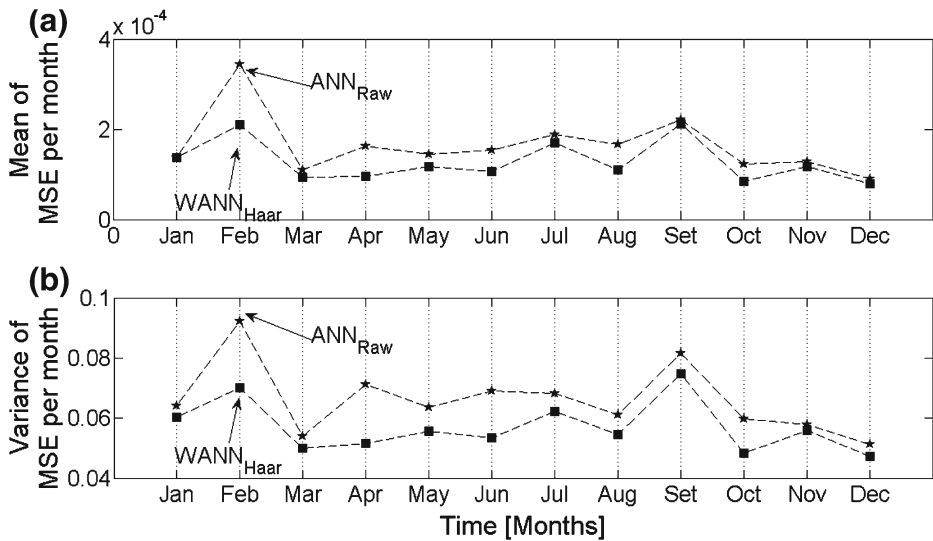
WANN<sub>Haar</sub> also provided more accurate results than ANN<sub>Raw</sub> for each and every month separately (Fig. 7). February and September were the most problematic months to be predicted; nevertheless *Haar* filtering increased forecasting accuracy for those months too. It was noted that relatively large errors are also coupled with large variance in model outputs.

Focusing on the best model only (i.e. WANN<sub>Haar</sub>), the forecasting error is normally distributed around a zero mean and in  $68.2\%$  of cases it is expected to be within  $\pm 0.038$  ( $10^6 \text{ m}^3/\text{month}$ ) from the mean. Visual observation suggested, and statistical analysis confirmed, that not all years can be predicted with the same accuracy (Fig. 8). For instance, the WANN<sub>Haar</sub> achieved best results in the forecasting of the year 2003, with a correlation coefficient between predicted and measured values equal to  $0.98$ ; on the other hand, in 2007, the same model achieved a correlation  $R=0.88$ .

**Table 3** Model-performances for 1 month-ahead forecasting

Model	Forward forecasting time-lag [months]	Indices of performances and efficiency				
		Correlation coefficient (R)	Root mean squared error (RMSE)	Fractional standard error. (FSE)	Nash Sutcliffe index (E)	Bias (B)
MLR	1	0.768	0.0592	0.033	0.561	0.996
ANN <sub>Raw</sub>	1	0.842	0.0458	0.026	0.708	1.000
WANN <sub>Haar</sub>	1	0.910	0.0361	0.020	0.818	1.000
WANN <sub>db2</sub>	1	0.896	0.0374	0.022	0.800	0.999
WANN <sub>db3</sub>	1	0.897	0.0387	0.022	0.797	0.999
WANN <sub>db4</sub>	1	0.832	0.0480	0.027	0.691	1.001
WANN <sub>db5</sub>	1	0.741	0.0583	0.032	0.548	0.999

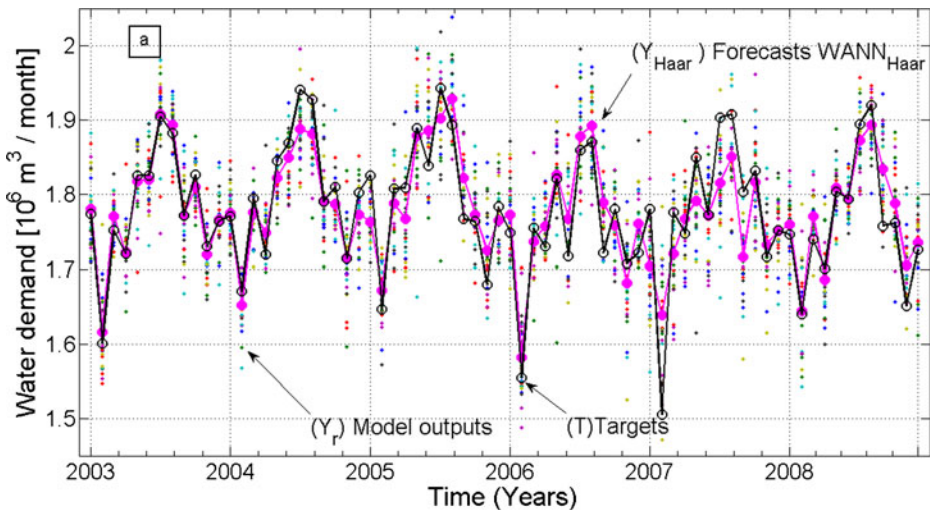




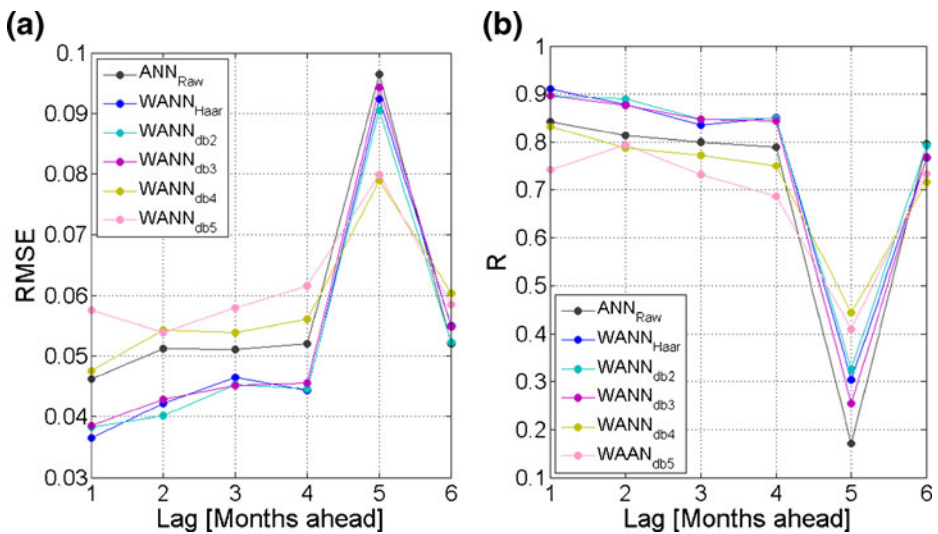
**Fig. 7** Comparison of WANN<sub>Haar</sub> versus ANN<sub>Raw</sub> on a month by month basis over the entire forecasting domain

For the multi-step ahead forecasting WANN<sub>Haar</sub>, as well as WANN<sub>db2</sub> and WANN<sub>db3</sub>, provided better results than non-coupled models at lag-time: 2, 3, 4 and 6 months (Fig. 9). A particular case is the 5th month ahead forecasting; at this time-lag: i) all models provided the largest errors and the lowest correlations; ii) in contrast to what had been previously observed, MLR, WANN<sub>db4</sub> and WANN<sub>db5</sub> returned the best performances. A table of results is presented in [Annex A](#).

When considering the variance of output generated by different iterations of the same model, it was observed that reducing variance in the input data-set (by means of denoising) resulted in increased variance in model outputs (Fig. 10). A particular case is *Haar* denoising; on the one



**Fig. 8** WANN<sub>Haar</sub> forecasts. Multi colored dots represent outputs of different experimental repetitions ( $Y_t$ ). Plain dots in violet represent model forecasts ( $Y_m = \text{mean of } Y_t$ ). Black circles represent model targets ( $T$ )



**Fig. 9** Error and correlation of model-output vs. targets at different time-lags

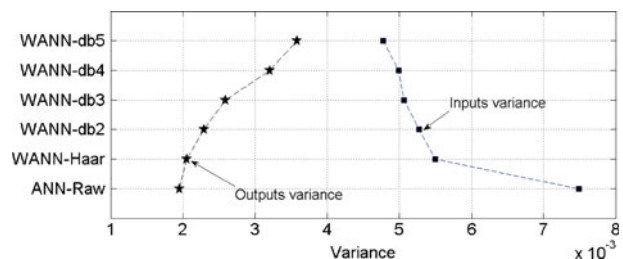
hand it did not affect ANN stability, while on the other hand, it drastically reduced input variance (Fig. 10), preserving and actually increasing ANN performances (Table 3). A different case is provided by *db5* denoising that induced an oversimplification of input patterns which resulted in both reduced ANN performances and increased ANN instability.

Results can be summarized as follows: *Haar*, *db2*, and *db3* filter-banks provided similar (and above average) performances for all lead times (with the exception of 5 months lead time); any extension in the future of the forecasting horizon corresponded to a decrease in forecasting accuracy; and lower variance in the input matrix resulted in higher instability of the outputs of the ANNs (with the exception of *Haar* filtering that provided high accuracy without affecting stability).

## 6 Discussion

ANNs do not require any a priori assumptions on the underlying process generating the data at hand. Nevertheless, in our case, the WANN<sub>dbN</sub> models are based on a fundamental assumption regarding the frequency domain of our dataset. In this respect, it is assumed that the original time series is likely to be affected by some type of noise either induced by intrinsic randomness in the studied phenomena, and/or imprecision of the water-meters.

**Fig. 10** Input-output variance of WANN models



Noise is usually concealed in the shortest frequencies of the time-series, while meaningful information is likely to be carried in the longest frequencies. Wavelets, orthogonal waves of finite duration, can be used for localized denoising of non-stationary signals. In our case, wavelet-denoising was applied by suppressing high-pass coefficients at the first scale of decomposition as this was found to provide the best results. It was observed that filter-banks of type *Haar*, *db2* and *db3* improved neural performances.

The correlation coefficients between measured and forecasted values (for a lead time of one month) of the non-denoised network was  $R=0.84$  while, the same network trained with *Haar*-denoised input had a correlation of  $R=0.91$ . The error over the entire testing domain followed suit, with  $RMSE = 0.045$  for the  $ANN_{Raw}$ , and  $0.36$  for the  $ANN_{Haar}$ . With respect to multi-step ahead forecasting (for 2, 3, 4 and 6 months ahead), *Haar*, *db2* and *db3* improved the performance of the model. Forecasting of the 5th month returned contradictory results. In all instances wavelet filter-banks of type *db4* and *db5* did not improve the forecasting task assigned to the ANN.

We suggest that removing part of the variance from the original time series simplified the generalization process in ANNs. When considering ANNs as a pattern recognition device, denoising reduces the heterogeneity of the patterns presented to the network, thus facilitating classification (i.e. *Haar*, *db2*, *db3*). On the other hand, an oversimplification of the training-set can create ambiguities resulting in lower forecasting performances as well as greater network instability (e.g. *db4*, *db5*). However, variance alone cannot entirely explain the impact of denoising on ANN performance.

The original impact of signal denoising has to be found in the shape and the length of the filter-banks. In this respect, according to our preliminary results it cannot be excluded that wavelet transforms provided the best denoising (reflected by improved accuracy of ANN forecasting) when the filters had their highest coefficients as close as possible to zero. Thus, in our case, causal filters seem to have provided the best results when they could emphasize the most correlated components of the parent time series.

## 7 Conclusions

In this study we addressed the forecasting of future water consumption of the municipality of Syracuse, Italy. The objective was to build an adaptive tool that when trained may be readily used for operational applications such as water budgeting. We also attempted to tackle a problematic gap in the theoretical foundations regarding the choice of wavelets to be used in conjunction with ANNs for water resources forecasting applications.

Experimental results showed that first level decomposition by means of wavelet-filters can improve generalization in ANNs. In particular, filter-banks of type *Haar*, *db2* and *db3* alternatively coupled with the same ANN, provided more accurate forecasting than the non-coupled ANN, the Multiple Linear Regression and the coupled models with wavelets of type *db4* and *db5* on the forecasting horizons from 1 to 4 and 6 months-ahead. Preliminary results suggest that in our case, attenuation of variance in the input vectors may be a significant driver of improvements in ANN forecasting. It is also observed that variance alone cannot entirely explain the impact of denoising on ANNs. Thus, it cannot be excluded that using filters that can emphasize the components of the parent time series with the highest autocorrelation may be a key issue in the coupled wavelet-neural network models. Additional work will be required for a complete understanding of these findings, and the use of larger and more diversified datasets may help to confirm the empirical results obtained in this study. It is suggested that future investigations should be oriented to the featuring of

tailor-made orthogonal wavelets, for time series denoising in an attempt to generate sparse neural-weight matrices in back propagation networks.

**Acknowledgments** This study was supported by the Norman Zavaloff Foundation whose help is greatly appreciated. This study was also partially funded by a NSERC Discovery Grant held by Jan Adamowski, and by the IWRM-SMART project of The Federal Ministry of Education and Research, Germany, and The Ministry of Science and Technology (MOST) of the State of Israel. The authors would also like to thank the anonymous reviewers for their valuable comments.

## ANNEX A

**Table 4** Model performances in multi ahead forecasting (from 2 to 6 months ahead)

Model	Forward forecasting time-lag [months]	Indices of performances and efficiency				
		Correlation coefficient (R)	Root mean squared error (RMSE)	Fractional standard error. (FSE)	Nash Sutcliffe index (E)	Bias (B)
MLR	2	0.802	0.056	0.031	-0.512	0.994
	3	0.800	0.055	0.031	-0.299	0.995
	4	0.799	0.055	0.031	-0.796	0.998
	5	0.803	0.054	0.030	-1.363	0.999
	6	0.832	0.050	0.028	-1.283	0.999
ANN <sub>Raw</sub>	2	0.813	0.051	0.029	-0.146	0.997
	3	0.799	0.051	0.029	-0.145	0.999
	4	0.789	0.052	0.029	-0.607	1.000
	5	0.171	0.096	0.054	-0.561	1.001
	6	0.796	0.052	0.029	-1.315	1.002
WANN <sub>Haar</sub>	2	0.879	0.042	0.024	-0.122	1.000
	3	0.835	0.047	0.026	-0.298	0.998
	4	0.851	0.044	0.025	-0.700	1.001
	5	0.303	0.092	0.052	-0.653	1.001
	6	0.768	0.055	0.031	-1.343	1.000
WANN <sub>db2</sub>	2	0.889	0.040	0.023	-0.114	1.000
	3	0.846	0.045	0.025	-0.224	1.001
	4	0.849	0.045	0.025	-0.715	0.999
	5	0.326	0.090	0.051	-0.667	1.000
	6	0.792	0.052	0.029	-1.218	1.000
WANN <sub>db3</sub>	2	0.875	0.043	0.024	-0.058	0.999
	3	0.847	0.045	0.025	-0.182	1.000
	4	0.843	0.046	0.026	-0.680	0.999
	5	0.254	0.094	0.053	-0.632	0.999
	6	0.769	0.055	0.031	-1.245	0.999
WANN <sub>db4</sub>	2	0.788	0.054	0.030	0.201	0.998
	3	0.772	0.054	0.030	-0.107	1.000
	4	0.750	0.056	0.031	-0.552	1.000
	5	0.445	0.079	0.044	-0.508	1.001

**Table 4** (continued)

Model	Forward forecasting time-lag [months]	Indices of performances and efficiency				
		Correlation coefficient (R)	Root mean squared error (RMSE)	Fractional standard error. (FSE)	Nash Sutcliffe index (E)	Bias (B)
WANN <sub>db5</sub>	6	0.716	0.060	0.034	−1.129	1.004
	2	0.794	0.054	0.030	0.168	0.997
	3	0.731	0.058	0.032	−0.145	0.998
	4	0.687	0.062	0.034	−0.471	0.999
	5	0.410	0.080	0.045	−0.430	1.000
	6	0.732	0.058	0.033	−1.174	0.999

## References

- Adamowski J (2008) Peak daily water demand forecast modeling using artificial neural networks. *J Water Resour Plan Manage* 134(2):119–128
- Adamowski J, Chan HF (2011) A wavelet neural network conjunction model for groundwater level forecasting. *J Hydrol* 407(1–4):28–40
- Adamowski J, Karapataki C (2010) Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. *J Hydrol Eng* 15:729–743
- Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J Hydrol* 390(1–2):85–91
- Adamowski K, Prokoph A, Adamowski J (2009) Development of a new method of wavelet aided trend detection and estimation. *Hydrol Process Spec Issue Can Geophys Union Hydrol Sect* 23:2686–2696
- Adamowski J, Fung Chan H, Prasher SO, Ozga-Zielinski B, Sliusarieva A (2012) Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour Res* 48:W01528. doi:10.1029/2010WR009945
- Adar E, Neuman SP, Woolhiser DA (1988) Estimation of spatial recharge distribution using environmental isotopes and hydrochemical data: mathematical model and application to synthetic data. *J Hydrol* 97:251–277
- An A, Shan N, Chan C, Cercone N, Ziarko W (1996) Discovering rules from data for water demand prediction: an enhanced rough-set approach. *Eng Appl Artif Intell* 9(6):645–653
- Blatter C (1998) Wavelet: a primer. Peters Natick, Massachusetts
- Bougadis J, Adamowski K, Diduch R (2005) Short-term municipal water demand forecasting. *Hydrol Process* 19:137–148
- Cannas B, Fanni A, See L, Sias G (2006) Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. *Phys Chem Earth* 31(18):1164–1171
- Chou CM (2011) A threshold based wavelet denoising method for hydrological data modeling. *Water Resour Manag* 25(7):1809–1830. doi:10.1007/s11269-011-9776-3
- Compo GP, Torrence C (1998) A practical guide to wavelet analysis. *Bull Am Meteorol Soc* 79:61–78
- Conover WJ (1980) Practical nonparametric statistics. Wiley, New York
- Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia
- Foufoula-Georgiou E, Kumar P (1994) Wavelets in geophysics. Academic, San Diego
- Grubbs JW (1994) Evaluation of ground-water flow and hydrologic budget for lake Five-O, a seepage lake in Northwestern Florida. U.S. Geological Survey, WRIR 94-4145
- Güldal V, Tongal H (2010) Comparison of recurrent neural network, adaptive neuro-fuzzy inference system and stochastic models in Eğirdir Lake level forecasting. *Water Resour Manag* 24:105–128
- Hagan MT, Demuth HB, Beale MH (1996) Neural network design. PWS Publishing, Boston
- Haykin S (1994) Neural networks, a comprehensive foundation. Macmillan College Publishing Company, New York
- Herrera M, Torgob L, Izquierdo J, Pérez-García R (2010) Predictive models for forecasting hourly urban water demand. *J Hydrol* 387(1–2):141–150

- Holder RL (1985) Multiple regression in hydrology. Institute of hydrology, Crowmarsh Gifford
- Holschneider M (1995) Wavelets: an analysis tool. Oxford Mathematical Monograph, Clarendon Press
- Issar AS, Zohar M (2009) Climate change impacts on the environment and civilization in the near East. In: Brauch HG, Spring ÚO, Grin J, Mesjasz C, Kameri-Mbote P, Behera NC, Chourou B, Krummenacher H (eds) Facing global environmental change. Springer, Berlin, pp 119–130
- Jain A, Ormsbee LE (2002) Short-term water demand forecasting modeling techniques: conventional versus AI. *J Am Water Works Assoc* 94(7):64–72
- Jowitt PW, Xu C (1992) Demand forecasting for water distribution systems. *Civ Eng Syst* 9:105–121
- Kisi O (2008) Stream flow forecasting using neuro-wavelet technique. *Hydrol Process* 22:4142–4152
- Kisi O (2009) Neural networks and wavelet conjunction model for intermittent stream flow forecasting. *J Hydrol Eng* 14(8):773–782
- Labat D (2005) Recent advances in wavelet analyses, part 1: a review of concepts. *J Hydrol* 314(1–4):275–288
- Labat D (2008) Wavelet analysis of the annual discharge records of the world's largest rivers. *Adv Water Resour* 31:109–117
- Maidment DR, Miaou SP (1986) Daily water use in nine cities. *Water Resour Res* 22(6):845–851
- Mallat S (1999) A wavelet tour of signal processing, 2nd edn. Academic, San Diego
- Meyer Y (1992) Wavelets and operators. Cambridge Studies in Advanced Mathematics, Cambridge University Press
- Msiza IS, Nelwamondo FV, Marwala T (2007) Artificial neural networks and support vector machines for water demand time series forecasting. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 7–10 October, Montreal, Canada, 638–643
- Nourani V, Komasi M, Mano A (2009) A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water Resour Manag* 23(14):2877–2894
- Oron G, Campos C, Gillerman L, Salgot M (1999) Wastewater treatment, renovation and reuse for agricultural irrigation in small communities. *Agr Water Manag* 38(3):223–234
- Partal T, Cigizoglu HK (2008) Estimation and forecasting of the daily suspended sediment data using wavelet-neural networks. *J Hydrol* 358(3–4):317–331
- Renaud O, Starck JL, Murtagh F (2005) Wavelet-Based Combined signal filtering and prediction. *IEEE T Syst Man Cy B* 35(6):1241–1251
- Rioul O, Vetterli M (1991) Wavelet and signal processing. *IEEE Signal Proc Mag* 8:14–38
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323:533–536
- Shirsath PB, Singh AK (2010) A comparative study of daily pan evaporation estimation using ANN, regression and climate based models. *Water Resour Manag* 24(8):1571–1581
- Smith J (1988) A model of daily municipal water use of short term forecasting. *Water Resour Res* 24(2):201–206
- Strang G, Nguyen T (1996) Wavelets and filter banks. Wellesley-Cambridge Press
- Wang W, Jin J, Li Y (2009) Prediction of inflow at three gorges dam in Yangtze River with wavelet network model. *Water Resour Manag* 23(13):2791–2803
- Widrow B, Lehr MA (1990) 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proc IEEE* 78(9):1415–1441
- Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50:159–175
- Zhang GP, Patuwo EB, Hu MY (1998) Forecasting with artificial neural network: the state of the art. *Int J Forecast* 14:35–62
- Zhang GP, Patuwo BE, Hu MY (2001) A simulation study of artificial neural networks for nonlinear time-series forecasting. *Comput Oper Res* 28(4):381–396
- Zhou H, Peng Y, Liang G (2007) The research of monthly discharge predictor-corrector model based on wavelet decomposition. *Water Resour Manag* 22(2):217–227



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.