NURW 1057182
10 July 2015
##########
CE: XX QA: KS
Coll:XX QC:XX

URBAN WATER JOURNAL, 2015
VOL. XX, NO. X, 1–10
http://dx.doi.org/10.1080/1573062X.2015.1057182

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Assessment of predictive uncertainty within the framework of water demand forecasting using the Model Conditional Processor (MCP)

S. Alvisi and M. Franchini

Department of Engineering, University of Ferrara, Via Saragat, Ferrara, Italy

ABSTRACT
This paper presents an application of the Model Conditional Processor (MCP), originally proposed by Todini (2008) within the hydrological framework, to assess the predictive uncertainty in water demand forecasting related to water distribution systems. The MCP enables us to assess the probability distribution of the future water demand conditional on the forecasts provided by two or more deterministic forecasting models. In the numerical application described here, where two years of hourly water demand data for a town in northern Italy are considered, two forecasting models are applied in order to forecast hourly water demands from 1 to 24 hours ahead: the first model has a modular structure comprising a periodic component which reflects the long-term effects and a persistence component which represents the short-term memory of the process; the latter is based on neural networks. The results highlight the effectiveness of the approach, provided that the data set used for the MCP parameterization is properly selected so as to be actually representative of the accuracy of the real-time water demand forecasting models.

ARTICLE HISTORY
Received 25 November 2014
Accepted 5 May 2015

KEYWORDS
water demand; forecasting; predictive uncertainty

1. Introduction

Accurate short-term water demand forecasting represents a fundamental prerequisite for efficient real-time management of water distribution systems, where the aim is to minimize the operating costs of pumping stations and contain network water losses, while ensuring that minimum levels of service and reliability are maintained. It is not surprising, therefore, that numerous short-term longitudinal (i.e. time series based) water demand forecasting models have been proposed in the scientific literature over the past twenty years. They generally fall into one of two large categories: models based on representing periodic patterns on different scales (e.g. Alvisi et al., 2007; Bakker et al., 2013; Shvartser et al., 1993; Zhou et al., 2002), and models based on data-driven techniques (e.g. Cutore et al., 2008; Ghiassi et al., 2008; Herrera et al., 2010; Jain et al., 2001).

However, all of this scientific production relies on "crisp" data and disregards the uncertainty connected to forecasting. One exception is the model proposed by Cutore et al. (2008), in which the uncertainty component connected to estimating the *parameters* is assessed using the SCEM-UA algorithm by Vrugt et al. (2003); however, it does not in itself represent the actual uncertainty of future demand.

In reality, as highlighted by Donkor et al. (2013) in a recent review of water demand forecasting models, *taking into account the uncertainty* within the framework of water demand forecasting is of fundamental importance, particularly when these models are used within the framework of water management procedures or to support decision-making (see also Bargiela, 1993; Hutton et al., 2014). Indeed, operation of water distribution systems requires a variety of decisions for determining, for example, water pumping schedules, reservoirs' rules and pressure control in the network (Bargiela, 1993). These decisions depend both on the information available at the moment in which the decision is taken and the attitude of the decision maker (Jamieson et al., 2007). For example, turning a pump on or off to charge a reservoir for the subsequent hour will depend, first of all, on the forecasted water consumptions, but also on the attitude of the decision-maker towards risk: s/he could be risk-prone when the level of damage (e.g. the tank run out or low pressures can occur at some network nodes) is low but s/he could be risk-averse when the level of damage is high. Very often, relying on deterministic forecasts, without any information about the probability of occurrence of future values, operators tend to stay on the side of caution (Jamieson et al., 2007). Indeed, remembering that the risk depends both on the damage and the probability of occurrence, in order to take rational decisions, it is necessary a) to define a utility function which depends on the damage, b) to estimate the forecasting uncertainty, and in particular the predictive density function of the value of interest conditional on all the information available at the time of forecasting, and then c) to minimize the expected value of the risk depending on a) and b).

Unfortunately, as mentioned earlier and to our knowledge, water demand forecasting models proposed in the scientific literature to date do not take into account the actual uncertainty of future demand, and this, given the previous considerations, certainly represents a serious limitation. In the field of hydrology, in contrast, over the past two decades numerous researchers

CONTACT S. Alvisi stefano.alvisi@unife.it

have investigated the problem of uncertainty tied (real-time) forecasting of hydrological variables such as hydrometric levels or discharges in a given river cross-section (see, for example, Alvisi & Franchini, 2011, 2012; Alvisi et al., 2012; Montanari & Grossi, 2008; Pelletier, 1987; Shrestha et al., 2009). Many different methods are employed and the terminology used is not always consistent and this has indirectly resulted in a "linguistic uncertainty" (Regan et al., 2003). In any case, these methods are generally structured so as to construct a chain (using probabilistic or non-probabilistic tools) between the uncertainty of a) the input data (for example, rainfall, upstream levels, etc.), b) the model parameters, c) the structure of the model and d) the output, as amply discussed in the papers by Shrestha and Solomatine (2006) and Montanari (2011). A distinctly different approach has been proposed, however, by Raftery, (1993), Krzysztofowicz (1999) and Todini (2008), where the uncertainty of the *real* future value is addressed by formally including the natural unpredictability of this value in a probability distribution that is conditional on the information available at the time of forecasting. This method, recently condensed by Todini (2008) into the Model Conditional Processor (MCP), is Bayesian in nature and can be used to estimate both the future expected value and its uncertainty on the basis of the forecasts (which summarize the information available at the time of forecasting) provided by one or more deterministic models. The recent applications of this model (Coccia & Todini, 2011) have demonstrated its validity and robustness.

Todini (2008) and Coccia and Todini (2011) also clearly demonstrate that the (global) uncertainty of the model, derived from the chain linking the input/model/parameters and output cannot be called *predictive uncertainty* but should rather be called *emulation uncertainty*. Indeed it is the predictive uncertainty that is actually of interest at the operational stage, since it regards the real future value of the variable being monitored (it can in fact be used to estimate water demand predictive density function to be used to minimize the expected value of the risk linked to a generic decision making process (Ostfeld, 2014) such as turning on/off a pump in order to increase the volume of water stored in a tank (Van Zyl et al., 2010) or issuing a water restriction alert during periods of low storage and inflow when prolonging water supply (Brennan et al., 2007; Chiew et al., 1998)). On the other hand, the emulation uncertainty has more value as a means of validating the model used in forecasting, since it shows the variability of its output relative to the actual future value.

Based on these latter considerations, which will be discussed in Section 2, and given the ability of the MCP to estimate predictive uncertainty, we thought it would be useful to analyze its potential in forecasting demand within a real water distribution system.

The rest of this paper is thus organized as follows. In Section 2 we focus on the meaning of predictive uncertainty (and how it differs from emulation uncertainty), which is the underlying concept of the MCP. Section 3 describes the structure of the MCP itself, or, rather, the manner in which it produces an estimate of future water demand and the associated uncertainty. As the MCP relies on forecasts provided by models which are external to the MCP itself, we present two water demand forecasting models having very different structural characteristics. Sections 4 and 5 describe the results obtained through the application of the MCP and the individual forecasting models in the case of a real water distribution system in northern Italy. Finally, in the last section we present some concluding considerations.

2. Predictive uncertainty

Predictive uncertainty is described by the probability distribution of the future (real) value of the *predictand* (in this case, the *average hourly water demand* in a water distribution system in the next hour or in one of the following 24 hours), which is conditional upon all the knowledge and information available at the time of forecasting (Krzysztofowicz, 1999; Raftery, 1993; Todini, 2008). At the operational stage, *all the knowledge gathered up to the time of forecasting is summarized in the predictions made by the (deterministic) model or models used*.

Predictive uncertainty must not be confused with *emulation uncertainty*. The difference between the two can be easily understood by introducing the joint probability distribution of the real value and the forecasted one(s) (Todini, 2008). Let us consider, for the sake of graphic simplicity, the case of a single forecasting model. Figure 1 provides an example of a possible joint (sampling) distribution of the real (q) and forecasted ($\hat{q}$) water demands. In general, the points ($\hat{q}$,q) form a cloud in which the degree of scatter will depend on the accuracy of the forecasting model: the higher the scatter, the lower the accuracy. In particular, by cutting the joint probability distribution by a preset real water demand q and re-normalizing it we obtain the conditional probability distribution of the forecasted water demand $\hat{q}$ given the real water demand q, i.e. $f(\hat{q}|q)$ (see Figure 1): this represents the *emulation uncertainty* of the model and basically quantifies how the values forecasted by the model are distributed when the water demand takes on a given value in reality. Therefore, it is a useful tool for assessing the forecasting model's performance. On the other hand, if we cut the joint probability distribution by a preset forecasted water demand value $\hat{q}$and re-normalize it, we will obtain the conditional probability distribution of the real water demand q given$\hat{q}$, i.e. $f(q|\hat{q})$ (see Figure 1): this represents the *true predictive uncertainty*, since it provides the probability distribution of the future real value given a certain value (a certain forecast) provided by the model. It is important to observe that it is this latter information, not the emulation uncertainty, which is of practical utility when it comes to using forecasts for water management purposes. It is precisely the *real* value of the water demand that will occur in the next hour (or in one of the next 24 hours) which may result in a system failure (e.g. emptying of a storage tank, etc.), not the value provided by the model.

3. The Model Conditional Processor (MCP)

The MCP is a Bayesian method proposed by Todini (2008) - drawing on the studies by Raftery (1993) and Krzysztofowicz (1999) - to estimate predictive uncertainty based on a set of historical observations and the corresponding values predicted by one or more forecasting models. Below we illustrate the MCP with reference to a case in which two forecasting models are used; we refer the reader to Todini (2013) for greater details about its application in the case of a generic number of forecasting models.

The method entails converting historical observations and the corresponding forecasted values into a normal space using

NURW 1057182
10 July 2015
##########
CE: XX QA: KS
Coll:XX QC:XX

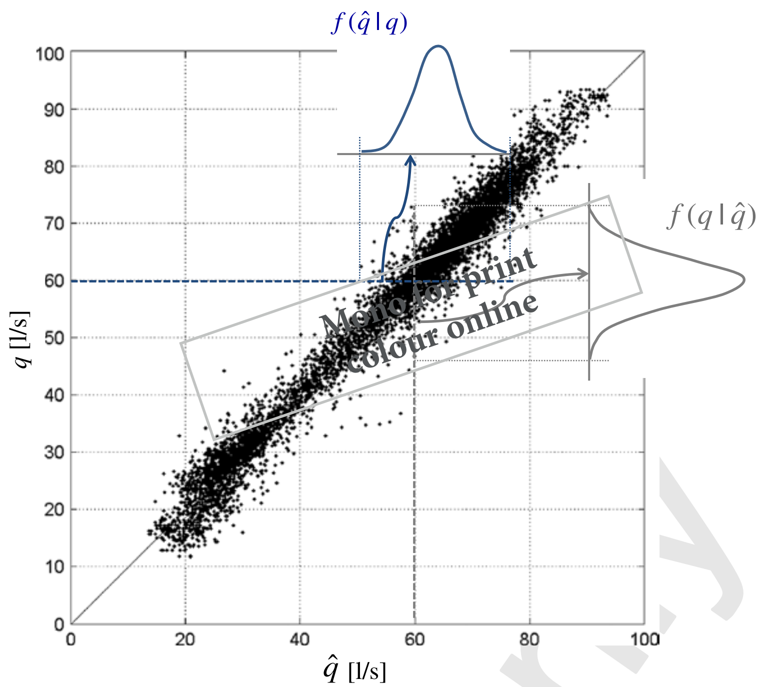

Figure 1. Example of joint sample frequencies of observed $\boldsymbol{q}$ and forecasted $\hat{q}$ water demands.

the Normal Quantile Transform (NQT) in order to arrive analytically at an estimate of the joint distribution of the real and forecasted values and hence at a conditional distribution of the *real values* given the *forecasted ones*. In greater detail, let q_t with $t = 1{:}n$ be the set of n observed values (for example hourly water demands across the network), $\hat{q}_{t|t-k\Delta t}$ the set of corresponding values forecasted $k\Delta$t time steps earlier (i.e. $k\Delta$t is the forecasting lead time) by means of a first forecasting model, and $\hat{\hat{q}}_{t|t-k\Delta t}$ the set of corresponding values forecasted $k\Delta$t time steps earlier by means of a second forecasting model. Via the NQT, we convert the set of n observed values q_t into a normal space, by arranging the data in ascending order and associating the corresponding cumulative sampling probability value with the i-th datum in the ordered vector using the Weibull plotting position $i/(n+1)$; each value q_t is thus associated with the corresponding value η_t, obtained from a standard normal distribution, the cumulative probability being equal. The same operation is repeated for the forecasted values $\hat{q}_{t|t-k\Delta t}$ and $\hat{\hat{q}}_{t|t-k\Delta t}$, so that we arrive at the values $\hat{\eta}_{t|t-k\Delta t}$ and $\hat{\hat{\eta}}_{t|t-k\Delta t}$, respectively. By virtue of the very nature of the NQT, η_t, $\hat{\eta}_{t|t-k\Delta t}$ and $\hat{\hat{\eta}}_{t|t-k\Delta t}$ are each distributed according to a normal distribution N(0,1). On the basis of these variables, we can therefore construct the joint distribution $\varphi_k\left(\eta_t, \hat{\eta}_{t|t-k\Delta t}, \hat{\hat{\eta}}_{t|t-k\Delta t}\right)$, assumed to be approximately normal (Todini, 2008), and having a mean:

$$\mu = \begin{bmatrix} \mu_\eta \\ \mu_{\hat{\eta}} \\ \mu_{\hat{\hat{\eta}}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{1}$$

and correlation matrix (Todini, 2013):

$$\mathrm{P} = \begin{bmatrix} 1 & \rho_{\eta\hat{\eta}} & \rho_{\eta\hat{\hat{\eta}}} \\ \rho_{\eta\hat{\eta}} & 1 & \rho_{\hat{\eta}\hat{\hat{\eta}}} \\ \rho_{\eta\hat{\hat{\eta}}} & \rho_{\hat{\eta}\hat{\hat{\eta}}} & 1 \end{bmatrix} \tag{2}$$

where $\rho_{\eta\hat{\eta}}$, $\rho_{\eta\hat{\hat{\eta}}}$ and $\rho_{\hat{\eta}\hat{\hat{\eta}}}$ represent the coefficients of correlation between η_t, $\hat{\eta}_{t|t-k\Delta t}$ and $\hat{\hat{\eta}}_{t|t-k\Delta t}$.

The conditional probability distribution $\varphi_k\left(\eta_t | \hat{\eta}_{t|t-k\Delta t}, \hat{\hat{\eta}}_{t|t-k\Delta t}\right)$ can therefore be derived from the joint distribution by exploiting the properties of multivariate normal distributions (see Mardia et al., 1979; Todini, 2008) in order to obtain a normal distribution $N\left(\mu_{\eta_t|\hat{\eta}_{t|t-k\Delta t},\hat{\hat{\eta}}_{t|t-k\Delta t}}, \sigma^2_{\eta_t|\hat{\eta}_{t|t-k\Delta t},\hat{\hat{\eta}}_{t|t-k\Delta t}}\right)$ whose conditional mean and variance are respectively:

$$\mu_{\eta_t|\hat{\eta}_{t|t-k\Delta t},\hat{\hat{\eta}}_{t|t-k\Delta t}} = \frac{\rho_{\eta\hat{\eta}}\left(\hat{\eta}_{t|t-k\Delta t} - \rho_{\hat{\eta}\hat{\hat{\eta}}}\hat{\hat{\eta}}_{t|t-k\Delta t}\right) + \rho_{\eta\hat{\hat{\eta}}}\left(\hat{\hat{\eta}}_{t|t-k\Delta t} - \rho_{\hat{\eta}\hat{\hat{\eta}}}\hat{\eta}_{t|t-k\Delta t}\right)}{1 - \rho^2_{\hat{\eta}\hat{\hat{\eta}}}} \tag{3}$$

$$\sigma^2_{\eta_t|\hat{\eta}_{t|t-k\Delta t},\hat{\hat{\eta}}_{t|t-k\Delta t}} = 1 - \frac{\rho^2_{\eta\hat{\eta}} - 2\rho_{\eta\hat{\eta}}\rho_{\eta\hat{\hat{\eta}}}\rho_{\hat{\eta}\hat{\hat{\eta}}} + \rho^2_{\eta\hat{\hat{\eta}}}}{1 - \rho^2_{\hat{\eta}\hat{\hat{\eta}}}} \tag{4}$$

Equations (3) and (4) thus provide, in the normal space, the expected value and the corresponding variance of the predictand, which are conditional on the forecasts $\hat{\eta}_{t|t-k\Delta t}$ and $\hat{\hat{\eta}}_{t|t-k\Delta t}$ (in the transformed plane) provided by the two models used. Again, in the normal space, it is easy to define an interval around the expected value of assigned probability (e.g. 95%).

For operational purposes, finally, the conditional probability distribution must be reconverted to the real space using the NQT. In order to carry out this operation, it is advisable to accompany the sampling distribution of real data obtained using the Weibull plotting position with curves that can better describe the tail of the distribution itself and in particular enable quantiles which have corresponding cumulative probabilities greater than $n/(n+1)$or less than $1/(n+1)$ to be converted into real space. For this purpose, as suggested by Coccia and Todini (2011), we can use the following curves:

$$p(q) = p_{low}\left[\frac{q - q_{\min}}{q(p_{low}) - q_{\min}}\right]^a \tag{5}$$

$$p(q) = 1 - (1 - p_{up})\left[\frac{q_{\max} - q}{q_{\max} - q(p_{up})}\right]^b \tag{6}$$

wherein p_{low} is a preset probability value below which the curve of Equation (5) is used and p_{up} is a probability value above which the curve of Equation (6) is used, $q(p_{low})$ and $q(p_{up})$ are the real values (water demands) corresponding to these probabilities derived from the sampling distribution, q_{min} and q_{max} are the minimum and maximum real values (of water demands), for which the cumulative probability is assumed to be equal to 0 and 1, respectively, and a and b are the parameters to be estimated so that the shape of the two curves of Equations (5) and (6) fits well, respectively, with the initial and terminal part of the sampling distribution of real data obtained by means of the Weibull plotting position. Further details about the parameterization of these curves are given in the numerical application. Obviously, in the operational phase Equations (5) and (6) are used in inverse form: given the probability associated with $\eta_t \in N(0,1)$, in turn identified by means of the MCP, we calculate the corresponding value q belonging to the original domain.

3.1 The forecasting models used by the MCP

In the previous section it was shown how the MCP can process the forecasts provided by two distinct forecasting models. Here below we describe the two models used by us in the context of water demand forecasting. The first one seeks to reproduce the periodic patterns of water demand taking into account short-term persistence components, and will be referred to hereinafter as Patt_for (Alvisi et al., 2007). The second deterministic forecasting model is based on neural networks and will be referred to hereinafter as ANN_for. It is worth highlighting the different structure of the two models, since the former seeks to enucleate and reproduce the periodic phenomena that are at the basis of the water demand time series, whereas the latter is a purely data-driven model, in which no heuristic knowledge about the phenomenon to be forecasted is introduced *a priori* into the structure. Indeed, it is worth noting that combination of models characterized by very different structures allows improving significantly the performances of the MCP. In fact, as shown by Coccia (2011), even within a completely different framework, i.e. river discharge forecasting, application of two distributed models with very similar structures (TETIS and TOPKAPI models) lead to just marginal gains in terms of forecast accuracy and reduction of the predictive uncertainty. Instead, the combination of a physically based model with a data driven model leads to greater improvements, and this is comprehensible considering the higher amount of information they provide with respect to two very similar models.

3.1.1 The Patt_for forecasting model

The Patt_for water demand forecasting model (Alvisi et al., 2007) is based on use of the periodic patterns present in the data time series in order to forecast the hourly water demands from 1 to 24 hours ahead. Starting from the observation that a seasonal and weekly cyclicity of daily water demands and a daily cyclicity of hourly water demands can typically be identified in the demand time series, the proposed model is structured in two modules. In the first module, which has the function of forecasting average daily water demand on the day (or days) comprising the 24 h for which we are forecasting the hourly water demands, account is taken of the seasonal and weekly cyclicity and of a short-term persistence. Analogously, in the second module, which, by using the output of the first module, provides the water demand forecast for the following 24 hours, account is taken of the daily cyclicity and of a short-term persistence.

In greater detail, the average daily water demand $q^{for}_{d,m}$ for the Julian date m is estimated by means of the following relationship:

$$q^{for}_{d,m} = \bar{q}^{s}_{d,m} + \overline{\Delta d}_{i,j} + \delta_m \tag{7}$$

where $\bar{q}^{s}_{d,m}$ is the long-term average daily water demand representing the seasonal periodic component, $\overline{\Delta d}_{i,j}$ is a daily correction factor which varies according to the season i (i = 1,…,4, winter, spring, summer, autumn) and day j of the week (j = 1,…,7, Monday,…, Sunday) and δ_m is a short-term daily persistence component modelled by means of an auto-regressive model AR(1) (Box et al., 1994).

The hourly module, like the daily module, is made up of periodic and persistence components. The mean hourly water demand q^{for}_t forecast for the generic hour t is given by:

$$q^{for}_t = q^{for}_{d,m} + \overline{\Delta h}_{i,j,h} + \varepsilon_t \tag{8}$$

where $q^{for}_{d,m}$ is the average daily water demand forecast by means of the daily module, $\overline{\Delta h}_{i,j,h}$ is the hourly pattern depending on the hour h (h = 1,…,24) of the day j of the week (j = 1,…,7, Monday,…, Sunday) and season i (i = 1,…,4, winter, spring, summer, autumn). Finally ε_t represents the hourly persistence component, which is obtained by means of a regressive model based on the residues/errors observed in the last instant of time and in the same hour to which the forecast pertains, but on the previous day.

All the patterns and parameters of the persistence models are estimated on the basis of the data for the case considered, with an eye to minimizing the mean square deviation over the calibration period.

3.1.2 The forecasting model ANN_for

The water demand forecasting model used here is based on a three-layer feedforward neural network featuring n_p inputs, n_h neurons in the hidden layer and n_o outputs (Alvisi, 2004; Alvisi & Franchini, 2003). In such a neural network, given the input vector $\mathbf{p}$ [n_p×1], the weight matrixes $\mathbf{U}$ [n_h×n_p] and $\mathbf{W}$ [n_o×n_h] and the bias vectors $\mathbf{a}$ [n_h×1] and $\mathbf{b}$ [n_o×1], the output vector $\mathbf{o}$ [n_o ×1] is given by:

$$\mathbf{o} = \mathbf{f_{out}}\left(\mathbf{W}\cdot\left(\mathbf{f_{hid}}(\mathbf{U}\cdot\mathbf{p} + \mathbf{a})\right) + \mathbf{b}\right) \tag{9}$$

where $\mathbf{f_{hid}}$ and $\mathbf{f_{out}}$ are the transfer function of the hidden and output layer respectively. The characterization of the ANN_for model entails defining the number n_p of neurons in the input layer, n_h in the hidden layer and n_o in the output layer, the trans-

NURW 1057182
10 July 2015
##########
CE: XX QA: KS
Coll:XX QC:XX

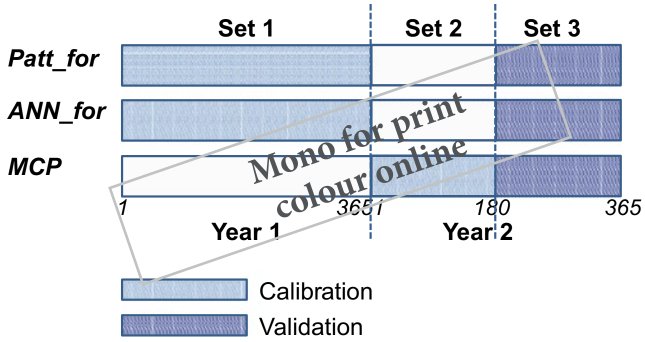

Figure 2. Data used for the calibration and validation of the individual deterministic forecasting models and for parameterization and validation of the MCP method.

fer functions $\mathbf{f}_{hid}$ and $\mathbf{f}_{out}$ and the weights **U** and **W** and biases **a** and **b**.

As regards the inputs, since the factors which most greatly influence hourly water demand are the periodic pattern over the day, on the one hand, and short-term fluctuations on the other, account is taken, respectively, of the 24 hourly water demands of the previous week corresponding to the 24 hours to be forecast and the last two observed hourly water demands. Thus, the number n_p of neurons in the input layer is set equal to 26. Indeed, test performed considering as inputs other variables such as climatic factors (average daily temperature, daily rainfall, etc.) showed that these information do not improve the accuracy of the forecasting model (Alvisi, 2004), and thus were not used.

As regards the number of hidden neurons, it is set equal to 72 in this case, based on previous experiences (the smallest number of neurons that can be used without excessively penalizing the model's performances, Hsu et al., 1995; Zealand et al., 1999).

As regards the number of neurons in the output layer, given that like the Patt_for model also the ANN_for model is specifically structured so as to provide water demand forecasts for the following 24 hours, n_o is set equal to 24.

The Tan-Sigmoid and linear transfer function are used in the hidden and output layers, respectively.

To avoid the problem of signal saturation (Hsu et al., 1995), the input data are normalized in the range [0.05:0.95]

Finally, the network is calibrated, and the weights and bias quantified accordingly using the Levemberg Marquardt algorithm (Hagan & Menhaj, 1994). In order to prevent overfitting and improve the robustness of the model, the technique of early stopping is used within the selected calibration period (ASCE, 2000; Demuth & Beale, 2000).

4. Case study

The MCP and individual forecasting models were applied to the water distribution system of Castelfranco Emilia (a municipality in the northern Italian province of Modena) with the aim of forecasting hourly demands [l/s] over a time horizon of 24 hours. The distribution network considered has an overall length of 160 km, serves around 23,000 inhabitants and is supplied through a tank, which is fed in turn by a well field located in proximity to the residential area of the town.

The measurements used to parameterize and validate the individual forecasting models and the MCP were taken on an hourly basis in the outlet pipe of the storage tank which feeds into the distribution network; the measurements thus represent the overall consumption of the network, including the water lost through leakage. It is important to observe that the overall water consumption across the network is precisely what needs to be taken into account for the real-time management of the system and proper planning of tank refills (Alvisi et al., 2007). The available data pertain to two years, 1998 and 2000, respectively years 1 and 2 in Figure 2. Both forecasting models were calibrated using the data of year 1 (set 1 of Figure 2). In particular, in the case of the Patt_for model, the data observed in year 1 were used to define the patterns which characterize the periodic components and to parameterize the AR(1) model and the regression that serve to represent the daily and hourly persistence components, respectively. In the case of the ANN_for model, the data observed in year 1 were equally divided between training and testing in order to calibrate the neural network using the early stopping technique.

Both models were validated using the data observed in the second half of year 2, i.e. set 3 of Figure 2

For the purpose of applying the MCP, the latter was parameterized using the data related to the first half of year 2 (set 2 in Figure 2) and validated using the data observed in the second half of year 2 (set 3) (see Figure 2). It is important to note that, given the presence of significant periodic components in water demand (Alvisi et al., 2007), the data were standardized prior to the application of the MCP: that is, the demand of the generic hour h of each day j was cleansed of the daily and seasonal fluctuation components connected to that hour h, represented by means of a Fourier series (Alvisi et al., 2007) of the data for that same hour, so as to convert it to a variable with a mean of 0 and variance of 1 by means of the following relation:

$$q^S_{h,j} = \frac{q_{h,j} - q^F_{h,j}}{\sigma_h} \tag{10}$$

where $q^S_{h,j}$ represents the standardized variable associated with the generic hour h of the day j (with $h = 1{:}24$), $q^F_{h,j}$ the flow rate of the hour h on the day j produced by means of the Fourier series, constructed in such a way that the differences $q_{h,j} - q^F_{h,j}$ have a mean 0 and standard deviation σ_h when computed over all the days j of the year.

Finally, with regard to the conversion of the variables by means of the NQT, the curves representing the tails of the sampling distribution obtained by means of the Weibull plotting

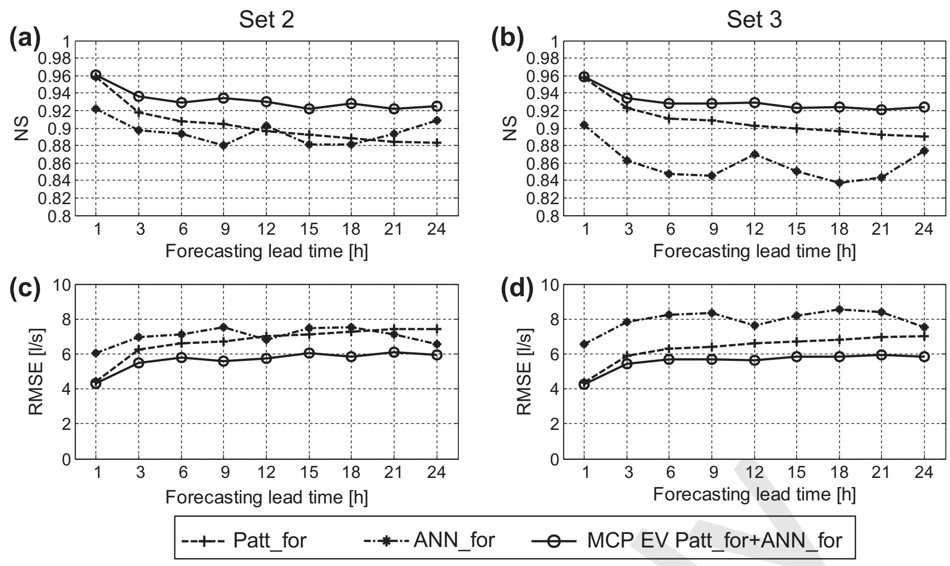

Figure 3. Comparison between the NS and RMSE coefficients for the Patt_for and ANN_for deterministic models and those related to the expected values (EV) provided by the MCP conditional on the Patt_for and ANN_for forecasts together, over the dataset used for the calibration (set 2) and validation of the MCP (set 3).

position (see Equations (5) and (6)) were defined assuming p_{low} (lower probability value, below which use is made of the curve of Equation (5)) to be equal to 0.01 and p_{up} (upper probability value, above which use is made of the curve of Equation (6)) to be equal to 0.99, and the lower and upper limits of these curves q_{min} and q_{max} (i.e. the minimum and maximum value for which it is assumed that the cumulative probability is equal to 0 and 1, respectively) to be equal, respectively, to a water demand which is null and one that double the maximum observed.

5. Analysis and discussion of the results

The results obtained using the MCP to forecast water demand under uncertainty were analyzed, with attention being focused on a) the accuracy and b) the uncertainty associated with the forecast provided.

5.1 Accuracy

The MCP was used to estimate the probability distribution of the real value conditional on the values forecasted by the two deterministic models. The mean of this distribution thus represents the expected value of the predictand which should be used for operational purposes in place of the deterministic forecasts provided by the individual models. The accuracy of this estimate was compared with that provided by the two individual models, Patt_for and ANN_for, considering the NS coefficient of Nash-Sutcliffe (Nash & Sutcliffe, 1970) and the root mean square error (RMSE), given respectively by:

$$\mathrm{NS} = 1 - \frac{\sum_{t=1}^{n} \left(q_t^{obs} - q_{t|t-k\Delta t}^{for} \right)^2}{\sum_{t=1}^{n} \left(q_t^{obs} - \mu_{q^{obs}} \right)^2} \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} \left(q_t^{obs} - q_{t|t-k\Delta t}^{for} \right)^2}{n}} \tag{12}$$

where q_t^{obs} (with $t = 1{:}n$) [l/s] represents the observed water demand, $\mu_{q^{obs}}$ [l/s] the mean of the observed demand, $q_{t|t-k\Delta t}^{for}$ [l/s] the water demand forecasted $k\Delta t$ time steps earlier (i.e. $k\Delta t$ is the forecasting lead time) by the Patt_for or ANN_for model or the expected value provided by the MCP, and n the number of data considered.

Figure 3 shows a comparison between the NS coefficient and RMSE provided by the Patt_for and ANN_for deterministic models versus the expected value provided by the MCP, in relation to the datasets used for the calibration (set 2) and validation (set 3) of the MCP itself. As may be observed, the Patt_for model delivered excellent performances as far as both set 2 and set 3 are concerned, with an NS coefficient ranging between a maximum value of about 0.96 for the shortest forecasting lead time (1 hour) and about 0.88 for the longest forecasting lead time (24 hours) Figure 3a and b); the ANN_for model likewise performed well, though the values of the NS coefficient were slightly lower, between 0.92 and 0.88 for set 2 (Figure 3a) and 0.91 and 0.84 for set 3 (Figure 3b). The expected value provided by the MCP enables us to obtain a forecast that is even more accurate than the ones provided by the aforementioned models, with values of the NS coefficient ranging between 0.96 and 0.92 for both set 2 and set 3. In particular, it is interesting to observe that the improvement in the NS coefficient compared to the values given by the individual deterministic models becomes more significant as the lead time lengthens. Similar considerations apply for the RMSE (see Figure 3c and d): the expected value provided by the MCP shows a reduced mean square error compared to the forecasts of both deterministic models. In general, therefore, the MCP can provide a more accurate forecast of future water demand than each of the two deterministic models alone since, by combining their forecasts, it exploits the information output by both of them. These results are in line with previous finding by Todini (2008), which however relates to the hydrological context of river discharge forecasting.

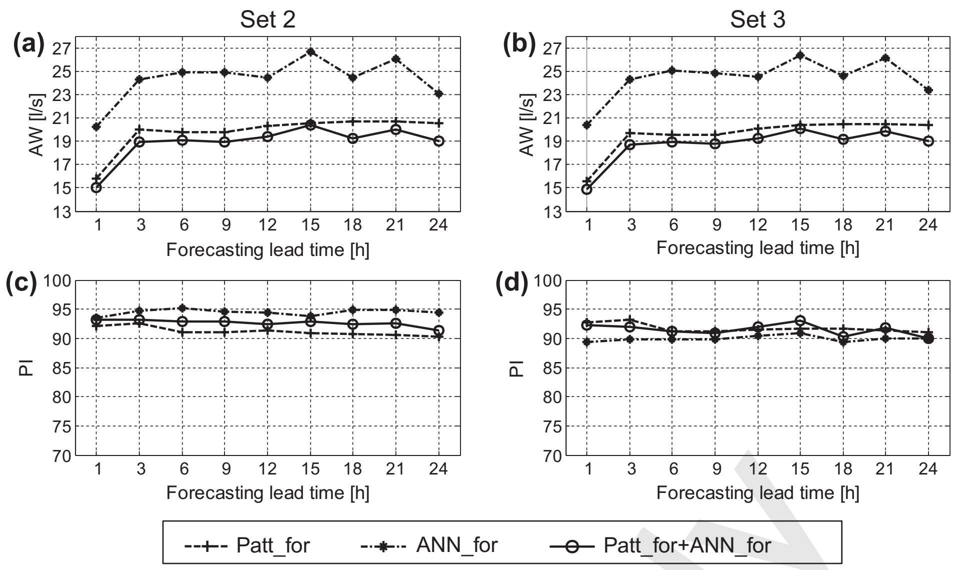

Figure 4. Average width (AW) and percentage of observed values actually included (PI) in the 95% uncertainty band conditional on the ANN_for forecasts, the Patt_for forecasts and the Patt_for and ANN_for forecasts together, in relation to the dataset used for the calibration (set 2) and validation of the MCP (set 3).

5.2 Uncertainty

Focusing our attention on the uncertainty associated with forecasting, we considered the average width (AW) of the 95% band and the percentage of observed values actually included in it (PI), given respectively by (see also Alvisi & Franchini, 2012; Xiong et al., 2009; Zhang et al., 2009):

$$AW = \frac{1}{n}\sum_{t=1}^{n}\left|q_t^{up} - q_t^{low}\right| \quad (13)$$

$$PI = \frac{1}{n}\sum_{t=1}^{n}\delta_t$$

$$\text{where} \quad \delta_t = \begin{cases} 1 & \text{if } q_t^{low} \le q_t^{obs} \le q_t^{up} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

q_t^{low} and q_t^{up} being respectively the lower end and upper end of the 95% band. In particular, these coefficients were evaluated based on the results provided by the MCP, considering the two deterministic forecasting models simultaneously. The predictive uncertainty associated with the MCP was also assessed considering only one deterministic model at a time, i.e. using the Hydrologic Uncertainty Processor (HUP) method (Krzysztofowicz, 1999) in the modified version developed by Todini (2008) (see also Todini, 2013).

Figure 4 shows the AW and PI obtained considering the two deterministic forecasting models individually, and in combination, for different lead times ranging between 1 and 24 hours.

It may be observed first of all that the bands obtained when we considered the forecasting models both individually and in combination do actually include a percentage of observed values close to 95% with respect to both set 2 and set 3 (Figure 4c and d); the width of these bands (Figure 4a and b) is greater for all forecasting lead times in the case of the ANN_for model, which thus displays not only a slightly lower accuracy, as observed previously (see Figure 3), but also a greater predictive uncertainty, with an average band width ranging between approximately 20 and 26 l/s depending on the lead time considered; the average width of the band we obtained considering the Patt_for model on its own is distinctly smaller, in the range of about 16 to 20 l/s, and it is interesting to observe that by combining the two models we obtain an uncertainty band that is slightly narrower still. In short, when formulated so as to use the information provided by two forecasting models, the MCP enables us to better combine that information, thereby reducing the uncertainty in water demand forecasting. This is also confirmed by Figure 5, which shows and compares the 95% uncertainty bands we obtained when considering the 1-hour-ahead forecasts for two days belonging to set 3. As may be observed, combining the forecasts provided by the two deterministic models using the MCP serves to reduce the width of the uncertainty band, which is even narrower than the one associated with the MCP itself when based on the individual forecasting models. To summarize, as was already observed in the hydrological context of river discharge forecasting (Todini, 2008), using the MCP enables us to combine the water demand forecasts provided by two (or more) deterministic models, thus drawing useful information from both models and arriving at a forecast that is more accurate than the ones obtained from either model on its own, and marked by less uncertainty. It is furthermore worth remembering that performances of the MCP are significantly influenced by the characteristics of the deterministic forecasting models more than the number of models themselves. Indeed, model characterized by very different structures, like those here considered, provide high amount of information which leads to great improvements in forecast (Coccia, 2011).

5.3 Analysis of the effects of the choice of dataset for parameterization of the MCP method

The previously analyzed results regarding forecasting accuracy and the estimation of predictive uncertainty were obtained by parameterizing the MCP using the data of set 2 (see Figure 2) and validating it with the data of set 3. It should be noted that

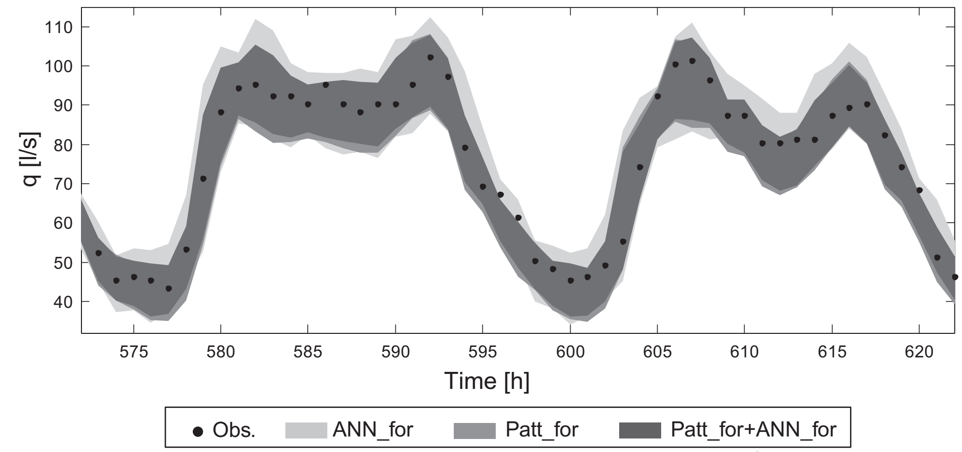

Figure 5. Comparison between the 95% Predictive Uncertainty band computed by using just one forecasting model (either Patt_for or ANN_for) and the two models together for 50 hours included in Set 3.

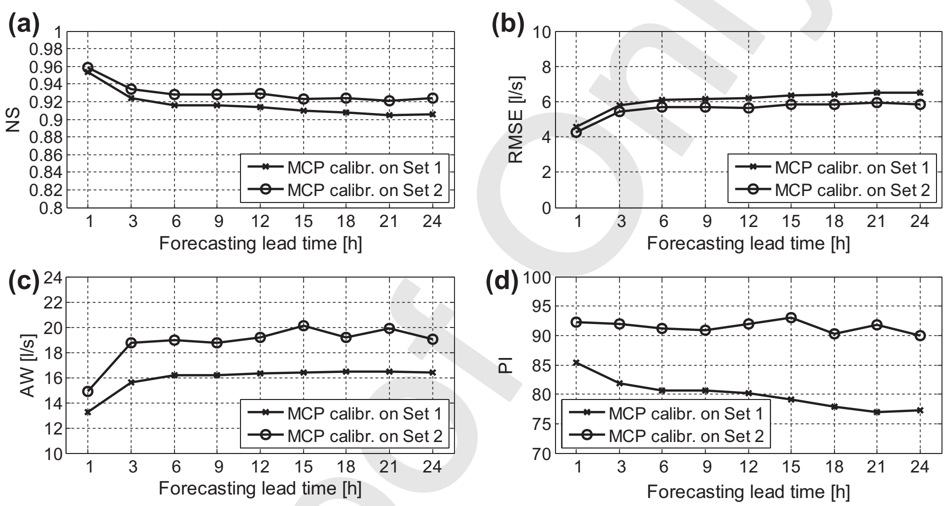

Figure 6. Set 3: Comparison of a) the NS, b) the RMSE of the expected value and c) the AW and d) the PI of the 95% uncertainty band provided by the MCP conditional on the Patt_for and ANN_for forecasts together, calibrated by using datasets 1 and 2.

the dataset used for parameterization, that is, for estimating the coefficients of correlation $\rho_{\eta\hat{\eta}}$, $\rho_{\eta\hat{\hat{\eta}}}$ and $\rho_{\hat{\eta}\hat{\hat{\eta}}}$ between the observed values (in the normal space) η and forecasted values $\hat{\eta}$ and $\hat{\hat{\eta}}$ (see Equations (3) and (4)), was not the same as the one used to calibrate the individual forecasting models (set 1 of Figure 2). This choice is justified by the fact that using a dataset which coincides with the one used for calibrating the individual forecasting models would have a negative impact on the performance of the MCP. To demonstrate this, in Figure 6. we show the results provided, during the validation step (set 3), by the MCP parameterized using both the data of set 1 and the data of set 2 (see Figure 2.). A change can be observed in the average width (AW) of the 95% uncertainty band and the corresponding percentage of observed values actually included (PI) (Figure 6c. and d): parameterizing the MCP with set 1 during the validation process results in an AW ranging between 13 and 16 l/s and a PI ranging between 75 and 85%; these values are distinctly lower than the ones obtained for the same set 3 when the MCP is parameterized using set 2.

This means that when the MCP is parameterized with the data of set 1, it provides a very narrow 95% uncertainty band which, in the validation phase, includes a distinctly lower percentage of observed values than the expected 95%. The result is understandable, considering that the individual water demand forecasting models provide a more accurate forecast in the calibration phase (set 1) than in the validation phase (set 3) (see Figure 7.). If the MCP is parameterized with the same dataset that was used to calibrate the individual forecasting models (set 1), their predictive uncertainty at the operational stage will not be captured and the 95% uncertainty band will be narrow(er). Its width will increase and it will become more reliable when the MCP is parameterized using a dataset (set 2) where the individual forecasting models operate outside their calibration range.

In short, the MCP, whose function is to quantify forecasting uncertainty, must be calibrated with the individual models used under operational conditions (outside calibration).

6. Conclusions

In the context examined here, the MCP enables us to combine the (short-term) demand forecasts of two or more models and provides a probability distribution of the real future demand

NURW 1057182
10 July 2015
##########
CE: XX QA: KS
Coll:XX QC:XX

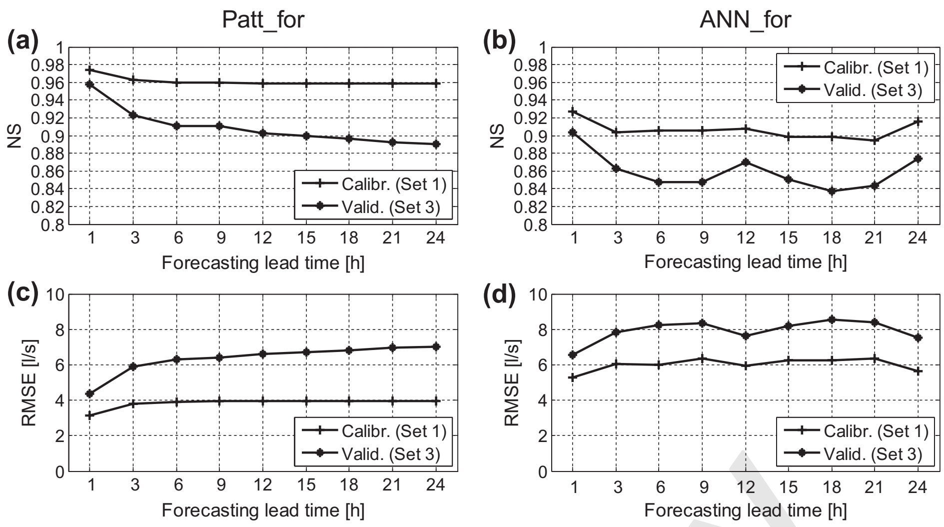

Figure 7. Comparison of the NS and RMSE coefficients computed for the calibration (set 1) and validation (set 3) sets of data for the Patt_for (a and c) and ANN_for (b and d) deterministic models.

which is conditional on the values predicted by the individual forecasting models. Based on this probability distribution, we can therefore estimate the expected value of future water demand and the associated predictive uncertainty (connected to its variance).

The method was applied to the demand data for a small town in northern Italy in order to forecast the water demands from 1 to 24 hours ahead using the forecasts provided by two models having a very different structure, one being based on a representation of the periodic components of demand, the other on the use of neural networks.

The results showed that the MCP, when appropriately parameterized, provides a future water demand forecast - represented by the expected value of the conditional probability distribution - which is more accurate than the ones provided by the individual forecasting models. It also reduces predictive uncertainty compared to that of the forecasting models considered individually.

In order to work correctly, the MCP must be parameterized using a dataset other than the one used to calibrate the individual forecasting models; only in this way will their predictive uncertainty be fully captured during the operational phase. Furthermore, within the framework of the NQT used to convert the observed and forecasted values into a normal space before applying the MCP, and to reconvert to the real space the outputs of the MCP, particular attention must be paid to extreme (high or low) values. In fact, in order to avoid inconsistent transformations, proper curve to describe the tail of the sampling value distribution should be used.

In conclusion, the MCP represents a tool which can be effectively applied for real-time water demand forecasting and can bring significant benefits for the management of water distribution systems, since it reveals the real uncertainty (probability density function) connected to the future level of demand which is the information necessary to estimate the risk connected to a decision.

Funding

The research has been funded by MIUR within the framework of PRIN 2012 project, titled "Tools and procedures for an advanced and sustainable management of water distribution systems", n. 20127PKJ4X and was carried out under the framework of Terra&Acqua Tech Laboratory, Axis I activity 1.1 of the POR FESR 2007 -2013, project funded by the Emilia-Romagna Regional Council (Italy) (http://fesr.regione.emiliaromagna.it/allegati/comunicazione/la-brochure-dei-tecnopoli).

References

Alvisi, S., & Franchini, M. (2003). La previsione in tempo reale delle richieste idriche nelle reti acquedottistiche. *Proceedings of La ricerca delle perdite e la gestione delle reti di acquedottoPerugia*, 26 September 2003 (In Italian).

Alvisi, S. (2004). *La previsione delle richieste idriche nel contesto della gestione in tempo reale delle reti di distribuzione di acqua potabile* (PhD Thesis). University of Ferrara, Italy (in Italian).

Alvisi, S., Franchini, M., & Marinelli, A. (2007). A short-term, pattern-based model for water demand forecasting. *Journal of Hydroinformatics*, *9*(1), 39–50.

Alvisi, S., & Franchini, M. (2011). Fuzzy neural networks for water level and discharge forecasting with uncertainty. *Environmental Modelling and Software*, 26(4), 523–537.

Alvisi, S., Creaco, E., & Franchini, M. (2012). Crisp discharge forecasts and grey uncertainty bands using data-driven models. *Hydrology Research*, *43*(5), 589–602.

Alvisi, S., & Franchini, M. (2012). Grey neural networks for river stage forecasting with uncertainty. *Journal of Physics and Chemistry of the Earth*, *42–44*, 108–118. AQ1

ASCE. (2000). Task Committee on the application of ANN in Hydrology: Artificial Neural Networks in Hydrology. I: Preliminary Concepts. *Journal of Hydrological Engineering*, *5*(2), 115–123.

Bakker, M., Vreeburg, J.H.G., van Schagen, K.M., & Rietveld, L.C. 2013. A fully adaptive forecasting model for short-term drinking water demand. *Environmental Modelling & Software*, *48*, 141–151.

Bargiela, A. (1993). Managing uncertainty in operational control of water distribution systems. In B. Coulbeck (Ed.). *Integrated Computer Applications* (pp. 353–363). Wiley.

Box, G.E.P., Jenkins, G.M., & Reinsel, G. C. (1994). *Time Series Analysis Forecasting and Control*. (3rd ed). Englewood Cliffs, NJ: Prentice-Hall.

Brennan, D., Tapsuwan, S., & Ingram, G. (2007). The welfare costs of urban outdoor water restrictions. *Australian Journal of Agricultural and Resource Economics*, *51*, 243–261.

Chiew F.H.S., Zhou, S.L., & McMahon, T.A. (1998). Use of seasonal streamflow forecasts in water resources management. *Journal of Hydrology*, *270*, 135–144.

Coccia, G., & Todini, E. (2011). Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrology and Earth System Sciences*, *15*, 3253–3274.

Coccia, G. (2011). *Analysis and Developments of Uncertainty Processors for Real Time Flood Forecasting* (PhD Thesis). University of Bologna, Italy. Retrieved
AQ2 from http://amsdottorato.unibo.it/3423/1/Tesi.pdf

Cutore P., Campisano A., Kapelan Z., Modica C. and Savic D. (2008). Probabilistic prediction of urban water consumption using the SCEM-UA algorithm. *Urban Water Journal*, 5(2), 125–132.

Demuth, H.B., & Beale, M. (2000). *Neural Network Toolbox for use with Matlab*. Natick: The Math Works.

Donkor, E., Mazzuchi, T., Soyer, R., & Roberson, J. (2012). Urban Water Demand Forecasting: A Review of Methods and Models. *Journal of Water Resources Planning and Management*. Doi: 10.1061/(ASCE)WR.1943-5452.0000314 (Sep. 13, 2012).

Ghiassi, M., Zimbra, D.K., & Saidane, H. (2008). Urban Water Demand Forecasting with a Dynamic Artificial Neural Network Model. *Journal of Water Resources Planning and Management*, *134*(2), 138–146.

Hagan, M.T., & Menhaj, M. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transaction on Neural Networks*, *5*(6), 989–993.

Herrera, M., Torgo, L., Izquierdo, J., & Perez-Garcia, R. (2010). Predictive models for forecasting hourly urban water demand. *Journal of Hydrology*, *387*(1–2), 141–150.

Hsu, K.L., Gupta, H.V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resource Research*, *31*(10), 2517–2530.

Hutton, C., Kapelan, Z., Vamvakeridou-Lyroudia, L., & Savic, D. (2014). Dealing with uncertainty in water distribution system models: a framework for real-time modeling and data assimilation. *Journal of Water Resources Planning and Management*, *140*(2), 169–183.

Jain, A., Varshney, A., & Joshi, U. (2001). Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. *Water Resources Management*, *15*(5), 299–321.

Jamieson, D.G., Shamir, U., Martinez, F., & Franchini, M. (2007). Conceptual design of a generic, real-time, near-optimal control system for water distribution networks. *Journal of Hydroinformatics*, *9*(1), 3–14.

Krzysztofowicz, R. (1999). Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resource Research*, *35*, 2739–2750.

Montanari, A., & Grossi, G. (2008). Estimating the uncertainty of hydrological
AQ3 forecasts: A statistical approach. *Water Resources Research*, *44*, W00B08. doi: 10.1029/2008WR006897

Montanari, A. (2011). Uncertainty of Hydrological Predictions. In Peter Wilderer (Ed.). *Treatise on Water Science* (vol. 2, pp. 459–478). Oxford: Academic Press.

Nash, J.E., & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models 1: a discussion of principles. *Journal of Hydrology*, *10*(3), 282–290.

Ostfeld, A. (2014). *Uncertainty and Risk Inclusions in Water Distribution Systems Management: Review and Challenges*. *Vulnerability, Uncertainty, and Risk, 1980–1985*.

Pelletier M.P., (1987). Uncertainties in the determination of river discharge: A literature review. *Canadian Journal of Civil Engineering*, *15*, 834–850.

Raftery, A.E. (1993). Bayesian model selection in structural equation models. In K.A. Bollen & J.S. Long (Eds.). *Testing Structural Equation Models* (pp. 163–180). Newbury Park, CA: Sage.

Regan, H.M., Akcakaya, H.R., Ferson, S., Root, K.V., Carroll, S., & Ginzburg, L.R. (2003). Treatments of uncertainty and variability in ecological risk assessment of single species populations. *Human Ecology Risk Assessment*, *9*, 889–906.

Shrestha, D.L., & Solomatine, D.P. (2006). Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks*, *19*(2), 225–235.

Shrestha, D.L., Kayastha, N., & Solomatine, D.P.. (2009). A novel approach to parameter uncertainty analysis of hydrological models using neural networks. *Hydrology and Earth System Sciences*, *13*, 1235–1248.

Shvartser, L., Shamir U., & Feldman M. (1993). Forecasting hourly water demands by pattern recognition approach. *Journal of Water Resources Planning and Management*, *119*(6), 611–627.

Todini, E. (2008). A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management*, *6*(2), 123–137.

Todini, E.. (2013). From HUP to MCP: Analogies and extended performances. *Journal of Hydrology*, *477*, 33–42.

Van Zyl, J.E., Piller, O., Legat, Y., & Walski, T. (2010). The risk of a municipal storage tank running dry due to user demands. *Proceedings of the Water Distribution System Analysis 2010 – WDSA2010*, 12–15 September, Tucson, AZ, USA.

Vrugt, J.A., Gupta, H.V., Bouten, W., & Sorooshian, S. (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, *39*(8), 1201.

Xiong, L., Wan, M., Wei, X., & O'Connor, K. (2009). Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation. *Hydrological Sciences Journal*, *54*(5), 852–871.

Zealand, C.M., Burn, D.H., & Simonovic, S.P. (1999). Short term streamflow forecasting using artificial neural networks. *Journal of Hydrology*, *214*, 32–48.

Zhang, X., Liang, F., Srinivasan, R., & Van Liew, M. (2009). Estimating uncertainty of streamflow simulation using Bayesian neural networks. *Water Resource Research*, *45*, W02403. doi: 10.1029/2008WR007030

Zhou, S., McMahon, T., Walton, A., & Lewis, J.. (2002). Forecasting operational demand for an urban water supply zone. *Journal of Hydrology*, *259*(1–4), 189–202.