Adaptable urban water demand prediction system

G. Banjac, M. Vašak and M. Baotić

ABSTRACT

In this work, identification of 24-hours-ahead water demand prediction model based on historical water demand data is considered. As part of the identification procedure, the input variable selection algorithm based on partial mutual information is implemented. It is shown that meteorological data on a daily basis are not relevant for the water demand prediction in the sense of partial mutual information for the analysed water distribution systems of the cities of Tavira, Algarve, Portugal and Evanton East, Scotland, UK. Water demand prediction system is modelled using artificial neural networks, which offer a great potential for the identification of complex dynamic systems. The adaptive tuning procedure of model parameters is also developed in order to enable the model to adapt to changes in the system. A significant improvement of the prediction ability of such a model in relation to the model with fixed parameters is shown when a certain trend is present in the water demand profile.

Key words | artificial neural networks, online parameters tuning, partial mutual information, water demand prediction

G. Banjac (corresponding author)
M. Vašak
M. Baotić
Faculty of Electrical Engineering and Computing,
University of Zagreb,
Unska 3,
HR-10000 Zagreb,
Croatia
E-mail: goran.banjac@fer.hr

INTRODUCTION

Improving the efficiency of water management is essential for overcoming the growing problem of water scarcity and droughts. Efficient water management implies efficient control of water distribution systems (WDSs), which is based on the availability and accuracy of data on variables that affect the WDS. The WDS control often relies on decision support systems (DSSs). An important element of such a DSS is the water demand prediction system (WDPS). Availability of accurate water demand prediction can lead to a decrease of overall energy and material consumption in the WDS, decrease of loads on elements of the WDS and increase of quality of water delivered to consumers (Bakker *et al.* 2013).

For the analysis in this paper, artificial neural networks (ANNs) are used for short-term water demand prediction with a time horizon of 24 hours. The procedure employed for a 1-hour-ahead water demand prediction is readily applicable for prediction of water demand in multiple hour-time-instants in the future based on the data available at the current time. Two approaches to such an extension are analysed in this paper: (i) based on dynamic employment of the 1-hour-ahead prediction model; and (ii) based on static models developed for each hour of the 24-hour prediction horizon.

MATERIAL AND METHODS

The overall water consumption, on the level of a water distribution network, consists of many different units. The pattern of water demand is a non-stationary stochastic time series that may include a trend in the mean, non-constant variance and discontinuities (Ghiassi *et al.* 2008). Research in this area shows that water demand includes non-linearities such that employment of ANNs in water demand prediction shows significantly better results compared with linear regression (Jain *et al.* 2001; Yalcinoz & Eminoglu 2005). Apart from the structure of a model, the performance of a WDPS depends on the model inputs as well.

Detailed methodology in ANN model development process is given in Maier *et al.* (2010) and is followed in this

paper. One of the most important steps in the model development process is input selection, where the vector of appropriate model inputs is determined. However, this step is usually not considered to be of great importance, and most of the input variables are determined heuristically, which can result in including too many or too few input variables (May *et al.* 2008). As a consequence of omitting one or more relevant input variables, the model will not be able to describe the whole dynamics and phenomena in the system. Probability of omitting relevant input variables is much higher for time series in which input candidates are not only different variables, but also their lagged values (unless dynamic ANNs are used), which significantly increases the number of potential input variables. Including too many input variables can be caused by poorly assessed relevance of an input variable or by existence of a redundancy among them, where some of the chosen variables contain some useful information, but are interdependent, so they contain redundant information. With an increase of input variables, a number of model parameters are also increased, which as a consequence leads to decreased speed and quality of the model calibration.

The linear correlation coefficient is a commonly adopted measure of dependence between variables. However, the underlying assumption of linearly structured dependence is contradictory to the model development of the non-linear system (May *et al.* 2008). A particular algorithm which is more appropriate for finding non-linear dependences, and which also considers the interdependences among variables, was developed by using the concept of partial mutual information (PMI) (Sharma 2000). In this paper, the original algorithm is adapted in the sense that PMI is estimated using the '*k*-th nearest neighbour' method described in Frenzel & Pompe (2007), which is more accurate and computationally less intensive than the original method. In addition, in relation to the original algorithm where the termination criterion is determined based on statistical significance of the estimated PMI, in this paper the predefined number of relevant inputs M is used as the termination criterion, i.e. the input variable selection (IVS) procedure creates a list of M most relevant inputs.

Multilayer perceptrons (MLPs) are the most common form of feed-forward ANN model architecture (Maier *et al.* 2010), and MLPs with one hidden layer are used in

this paper for modelling the demand. Feed-forward back-propagation method with Levenberg-Marquardt technique is used as the ANN training procedure because of its advantages in relation to the other gradient-descent techniques (Cigizoglu & Kişi 2005). The available data set for training the ANN is divided into three subsets: 'training data' (70%) are used for calculating the gradient and updating the model parameters; 'testing data' (15%) are used in the stopping criterion of the training process (cross-validation); and 'validation data' (15%) are used to determine the optimal number of inputs (out of M) and optimal number of hidden neurons. A set of ANNs with different number of inputs, which are chosen as first m elements from a list of M most relevant inputs, $m \leq M$, and different number of hidden neurons is trained – the optimal ANN is chosen as one with the lowest mean squared error (MSE) on the validation data. MSE is defined as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2, \tag{1}$$

where n is the number of data samples for which MSE is calculated, Y_i is the actual output of the ith sample and $\hat{Y}_i$ is the model output for the ith sample.

The above-described procedure can be used to identify a 1-hour-ahead demand prediction model. For identifying a multiple-hours-ahead demand prediction model, we propose two approaches: (i) based on dynamic employment of the 1-hour-ahead prediction model; and (ii) based on static models developed for each hour of the 24-hour prediction horizon. These approaches are shown in Figure 1.

It is often the case that historical data used for calibrating the demand model do not cover the complete set of possible input–output vectors, or that water demand that occurred in the past differs from demand for the coming period due to factors which were not considered or did not have a significant impact on demand during model calibration. For robust operation of the WDPS, the model should be able to adapt to possible changes in the system. Adaptive structure of the WDPS is shown in Figure 2. The system is composed of two parts: 'offline' and 'online'. In the offline part, historical data are used for obtaining the initial demand model. The online part of the WDPS uses the initial demand model for estimating the newly coming
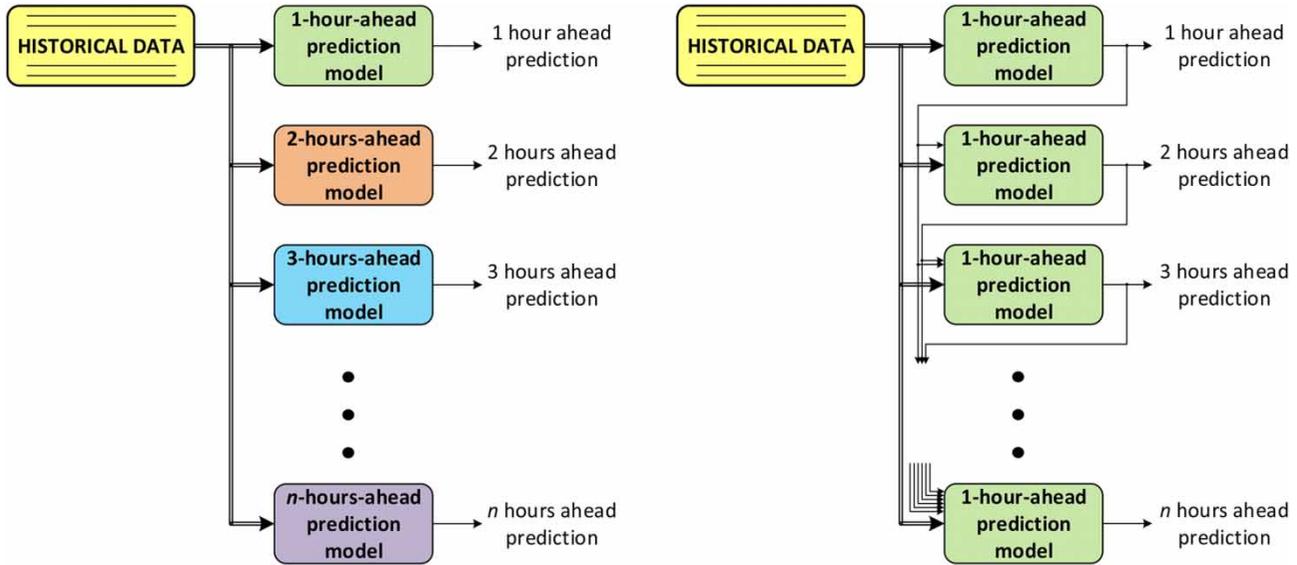
Figure 1 | Two approaches in identification of multiple-hours-ahead demand prediction model: static approach which uses a separate model for each hour (left) and dynamic approach which uses a single model for the whole prediction horizon (right).
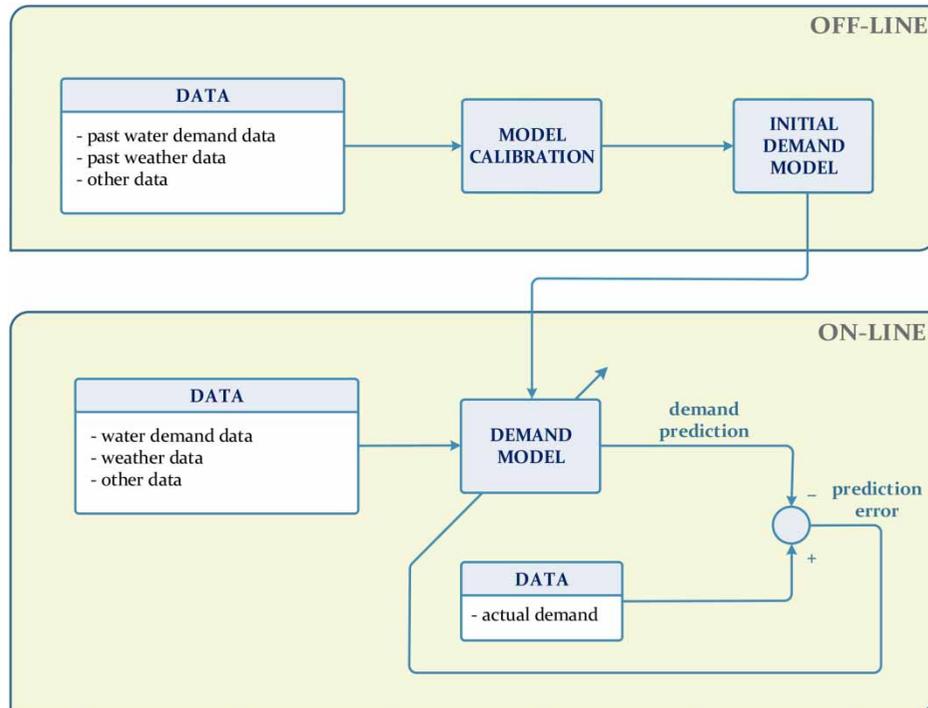
Figure 2 | Adaptive structure of the WDPS — an overview of the functionality.

water demand. When the actual demand data are available, they are compared with the corresponding demand prediction, which results in the prediction error for a certain time instant. Model parameters are then tuned such that the prediction error is decreased. The presented procedure of using the feedback information on prediction accuracy for model parameters tuning introduces an adaptation ability to the WDPS. There are a number of methods

suggested in the literature for the so-called 'recursive' ANN learning. Some of them are based on the recursive approximation of typical gradient methods (Ngia & Sjoberg 2000), while some of them are based on the methodology for dynamic system state estimation (Rivals & Personnaz 1998; Van der Merwe & Wan 2001; Zhan & Wan 2006; Wu & Wang 2012). ANN models with a relatively large number of inputs and nodes in the hidden layer result in a large number of parameters, and applying methods based on the latter methodology becomes intractable due to the numerical stability issues (Padilla & Rowley 2010). In this paper, the recursive gradient-descent method with momentum term is used for online tuning of model parameters – it is numerically stable even for ANNs with a large number of parameters, but still more flexible than the simple recursive gradient-descent method. ANN parameters Θ are updated based on the following relation:

$$\Delta\Theta(k) = -\alpha\nabla\mathfrak{I}(\Theta(k)) + \gamma_{\mathrm{m}}\Delta\Theta(k-1), \tag{2}$$

where $\Delta\Theta(k) = \Theta(k+1) - \Theta(k)$, α is the learning coefficient, $\nabla\mathfrak{I}(\Theta(k))$ is the gradient of local criterion function on the corresponding data set, and γ_{m} is the non-negative momentum term which speeds up the learning convergence while attenuating the parasitic oscillations.

RESULTS AND DISCUSSION

The steps followed in this paper for ANN modelling are described in Maier *et al.* (2010). In May *et al.* (2008), the IVS procedure based on the PMI is described and the advantages of using this measure are discussed. In this paper, PMI is calculated based on the algorithm presented in Frenzel & Pompe (2007), which is more accurate, computationally less intensive and less sensitive to parameter selection compared with the original algorithm. To the best of the authors' knowledge, online tuning of the water demand prediction has not been reported before.

The WDPS is developed for the selected district metering area in the city of Tavira, Algarve, Portugal, using MATLAB Neural Network Toolbox. Hourly water demand data, daily meteorological data and a set of time indices, which provide information about time of day, day in week

etc., are used as the candidates for model inputs. Historical data for the period from January 2011 until December 2012 are used for training the initial demand model. Using the developed IVS procedure, a set of $M = 20$ variables are chosen, as the most relevant for the 1-hour-ahead demand prediction in the sense of PMI, in the following order: $d(t-168)$, $d(t-24)$, $d(t-1)$, $t_{c,\mathrm{D}} = \cos(2\pi t_{\mathrm{D}}/24)$, $t_{s,\mathrm{D}} = \sin(2\pi t_{\mathrm{D}}/24)$, $d(t-48)$, $d(t-167)$, $t_{s,\mathrm{W}} = \sin(2\pi t_{\mathrm{W}}/168)$, $t_{c,\mathrm{W}} = \cos(2\pi t_{\mathrm{W}}/168)$, $d(t-23)$, $d(t-169)$, $d(t-47)$, $d(t-25)$, $d(t-49)$, $d(t-2)$, $d(t-170)$, $d(t-50)$, $d(t-26)$, $d(t-12)$ and $d(t-11)$, where $d(t-k)$ denotes demand k hours before the hour for which the prediction is to be determined, t_{D} is hour of day and t_{W} is hour in week from the last Monday midnight. The model with 20 inputs was chosen as the optimal one, and MLP with one hidden layer that consists of 25 neurons showed the best performance on the independent validation data (Table 1). Figure 3 depicts the regression plots for the chosen 1-hour-ahead prediction model on calibration (union of training and testing data) and validation data. Notation R stands for the regression value.

The procedure described above was also used for the identification of k-hours-ahead prediction models, $k \in \{2, \ldots, 24\}$. Both static and dynamic approaches for 24-hours-ahead demand prediction were tested, and it was shown that they have quite similar prediction abilities. However, from the computational point of view, applying the dynamic approach is much more efficient, since only one model needs to be identified and a significantly smaller set

Table 1 | Mean squared errors (MSE) of the obtained 1-hour-ahead demand prediction models on the calibration and validation data sets

Model	Calibration data MSE	Validation data MSE
Linear	91.7873	89.6176
MLP(15)	67.6150	75.6435
MLP(20)	64.0207	76.0082
MLP(25)	66.2181	74.2241
MLP(30)	66.0955	77.8976
MLP(35)	66.0251	75.3533
MLP(40)	64.9425	75.0948
MLP(45)	65.9636	77.0206
MLP(50)	65.0910	75.8998

MLP(*i*) stands for the multilayer perceptron neural network with *i* neurons in the hidden layer.
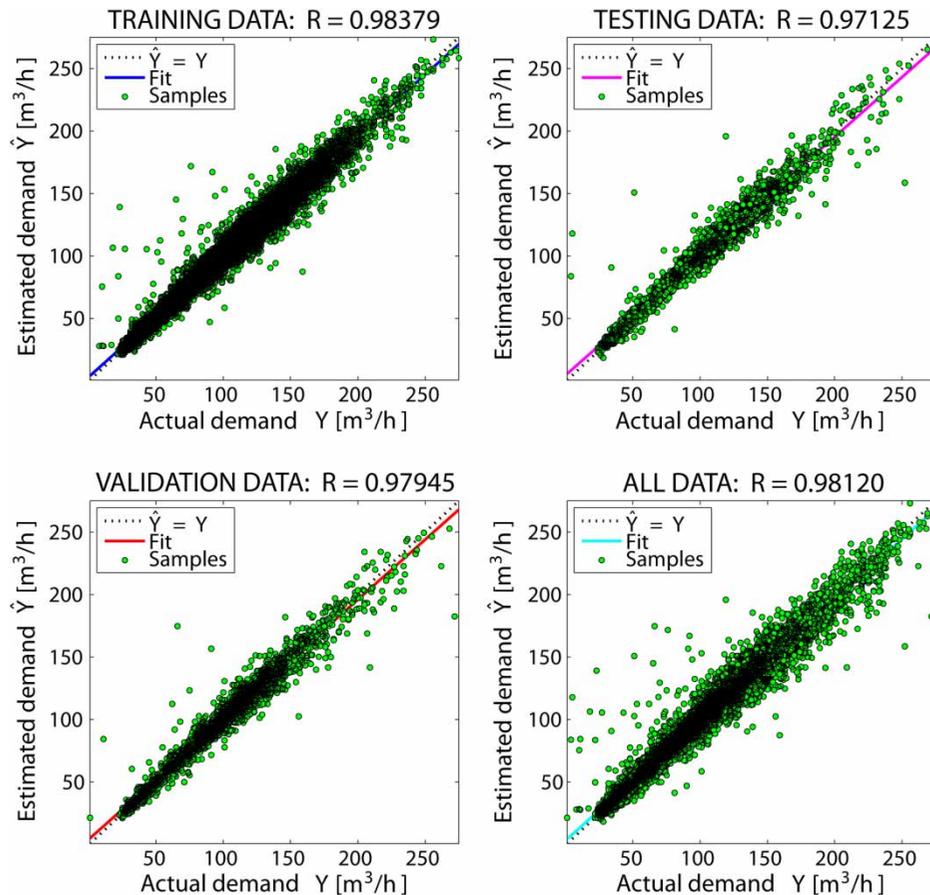
Figure 3 | Linear regression for actual and estimated water demand on different data sets for the 1-hour-ahead demand prediction model for the WDS of Tavira.

of parameters have to be tuned if model adaptation is used. In addition, extension of the prediction model to a larger prediction horizon using dynamic approach is trivial, while for the static approach additional models should be developed.

Finally, the recursive gradient-descent method with momentum term is used for the adaptive ANN parameters tuning, which is performed once a day, when demand data for the day before are available. Since some of the model inputs are time indices indicating a day in the week, using the information on prediction error for a single day in adaptation may cause the model to adapt to the specific day. Therefore, we use information on prediction error of the last 7 days for the model update. To test the performance of the online tuning procedure when a trend in demand is present, linear and sinusoidal signals are added to the original demand data for the period from January 2013 until December 2013. The linear signal is chosen such that its

value at the beginning of the 1-year period is equal to zero, and at the end of the period is equal to the mean of the original demand. The sinusoidal signal is chosen such that its amplitude is equal to the mean of the original demand and its period is equal to 2 years. Table 2 shows a significant improvement of the online model prediction ability in the sense of MSE compared with the offline model

Table 2 | Performance of the offline and online 24-hours-ahead water demand prediction models evaluated on validation data sets for the WDS of Tavira

Validation data	Offline model MSE	Offline model MAPE (%)	Online model MSE	Online model MAPE (%)
Original	98.6620	8.7325	100.0479	8.8047
Original + linear	214.9956	6.8601	146.1174	5.9160
Original + sine	658.8104	8.9333	416.3226	8.2597
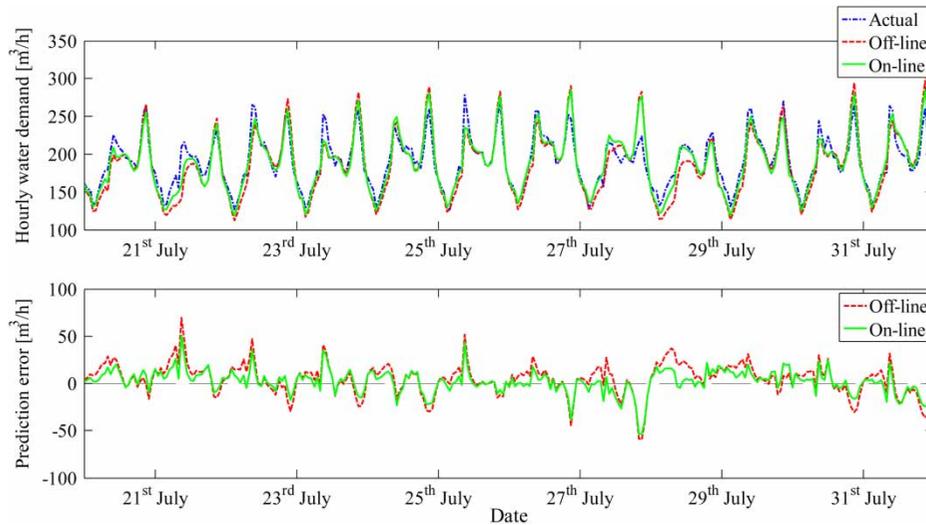
Figure 4 | Actual and day-ahead predicted water demand curves with artificially added linear signal for the WDS of Tavira in 2013 (prediction is made each day at midnight).

when there is a trend in demand. To get a better insight into the model performance, mean absolute error (MAPE) is also shown in the table. MAPE is defined as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \cdot 100\%. \tag{3}$$

Figure 4 shows that the online model tracks the actual demand curve better than the offline model does when a linear trend is added to the original demand.

To show applicability of the presented approach to other regions, the same analysis was performed for the data set of Evanton East, Scotland, UK. For brevity of the presentation, we only show performance of the offline and online 24-hour-ahead water demand prediction models in Table 3. The experiment was run in the same conditions as for the

Table 3 | Performance of the offline and online 24-hours-ahead water demand prediction models evaluated on validation data sets for the WDS of Evanton East

Validation data	Offline model MSE	Offline model MAPE (%)	Online model MSE	Online model MAPE (%)
Original	0.03088	4.821	0.03003	4.803
Original + linear	0.04096	4.882	0.03656	4.713
Original + sine	0.07098	5.463	0.05758	4.980

Tavira site. Results presented in Table 3 show even better performance of the offline model in terms of the MAPE than those in Table 2. Improvement of the prediction accuracy when using the online tuning procedure follows the same trend as in the case of the Tavira site. Note that MSE is not a relative statistic measure (see (1)), i.e. it is dependent on the units used for representing demand. Thus, it is not appropriate for comparing model performance for the two sites.

CONCLUSIONS

This paper considers the identification of the water demand prediction model with a time horizon of 24 hours using two approaches, static and dynamic. It was shown that both analysed approaches have quite similar predicting abilities. However, from the computational point of view, applying the dynamic approach is much more efficient. Moreover, extension of the prediction model to a larger prediction horizon using the dynamic approach is trivial, while this is not the case for the static approach.

It was shown that for the analysed WDSs of the Tavira and Evanton East sites, daily meteorological variables are not relevant, in the sense of PMI, for describing the demand prediction, i.e. past demand and variables describing time of day and day in week can satisfactorily model 1-hour-ahead water demand prediction model.

Finally, the experimental results show that using the online tuning procedure of model parameters can lead to a significant improvement of the model prediction ability, in the sense of MSE, when a certain trend exists in the water demand profile. Also, the applicability of the approach for the WDPS modelling was confirmed for two distinct sites.

ACKNOWLEDGEMENTS

REFERENCES

Bakker, M., Vreeburg, J. H. G., Palmen, L. J., Sperber, V., Bakker, G. & Rietvald, L. C. 2013 Better water quality and higher energy efficiency by using model predictive flow control at water supply systems. *J. Water Supply Res. Technol. AQUA* **62** (1), 1–13.

Cigizoglu, H. K. & Kişi, Ö. 2005 Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data. *Nord. Hydrol.* **36** (1), 49–64.

Frenzel, S. & Pompe, B. 2007 Partial mutual information for coupling analysis of multivariate time series. *Phys. Rev. Lett.* **99** (20), 204101.

Ghiassi, M., Zimbra, D. K. & Saidane, H. 2008 Urban water demand forecasting with a dynamic artificial neural network model. *J. Water Resour. Plann. Manage.* **134** (2), 138–146.

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

Maier, H. R., Jain, A., Dandy, G. C. & Sudheer, K. P. 2010 Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Env. Model. Softw.* **25** (8), 891–909.

May, R. J., Maier, H. R., Dandy, G. C. & Fernando, T. M. K. G. 2008 Non-linear variable selection for artificial neural networks using partial mutual information. *Env. Model. Softw.* **23** (10–11), 1312–1326.

Ngia, L. S. H. & Sjoberg, J. 2000 Efficient training of neural nets for nonlinear adaptive filtering using a recursive Levenberg-Marquardt algorithm. *IEEE Trans. Signal Process.* **48** (7), 1915–1927.

Padilla, L. E. & Rowley, C. W. 2010 An adaptive-covariance-rank algorithm for the unscented Kalman filter. In: *49th IEEE Conference on Decision and Control*, pp. 1324–1329. http://dx.doi.org/10.1109/CDC.2010.5717549.

Rivals, I. & Personnaz, L. 1998 A recursive algorithm based on the extended Kalman filter for the training of feedforward neural models. *Neurocomputing* **20** (1–3), 279–294.

Sharma, A. 2000 Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 – A strategy for system predictor identification. *J. Hydrol.* **239** (1–4), 232–239.

Van der Merwe, R. & Wan, E. A. 2001 The square-root unscented Kalman filter for state and parameter-estimation. *IEEE Int. Conf. Acoust. Speech Signal Process. Proc.* **6**, 3461–3464. http://dx.doi.org/10.1109/ICASSP.2001.940586.

Wu, X. & Wang, Y. 2012 Extended and unscented Kalman filtering based feedforward neural networks for time series prediction. *Appl. Math. Model.* **36** (3), 1123–1131.

Yalcinoz, T. & Eminoglu, U. 2005 Short term and medium term power distribution load forecasting by neural networks. *Energy Convers. Manage.* **46** (9–10), 1393–1405.

Zhan, R. & Wan, J. 2006 Neural network-aided adaptive unscented Kalman filter for nonlinear state estimation. *IEEE Signal Process. Lett.* **13** (7), 445–448.