

# Water Demand Modelling Using Independent Component Regression Technique

Md Mahmudul Haque<sup>1</sup> · Amaury de Souza<sup>2</sup> ·  
Ataur Rahman<sup>1,3</sup>

Received: 27 October 2015 / Accepted: 12 October 2016 /  
Published online: 20 October 2016  
© Springer Science+Business Media Dordrecht 2016

**Abstract** Water demand modelling is an active field of research. The modelling and forecasting tools are useful to get the estimation of forecasted water demand for different forecast horizons (e.g. 1 h to 10 years) in order to achieve more efficient and sustainable water resources management systems. However, modelling and forecasting of accurate water demand are challenging and difficult tasks. Several issues make the demand forecasting challenging such as the nature and quality of available data, numerous water demand variables, diversity in forecast horizons and geographical differences in modelling catchments. These issues have motivated a number of studies to be conducted to produce better water demand modelling and forecasting tools in order to improve forecast reliability. A variety of techniques have been adopted in water demand forecasting, however, application of independent component regression (ICR) technique has not been investigated yet. Hence, this study explores, for the first time, the use of the ICR technique in medium term urban water demand forecasting. This uses data from the city of Aquidauana, Brazil. It also compares the performance of the developed ICR model with two other commonly modeling methods, principal component regression and multiple linear regression models. It has been found that ICR model perform better than the other two models in modelling water demand with a higher performance indices (i.e.  $R^2$ , RMSE, NSE and MARE) for the independent validation period. The results indicate that the ICR technique has the potential to develop water demand models successfully. The methodology adopted in this paper can be applied to other countries to develop water demand forecasting model.

---

✉ Ataur Rahman  
a.rahman@uws.edu.au

<sup>1</sup> Institute for Infrastructure Engineering, Western Sydney University, Penrith, NSW 2751, Australia

<sup>2</sup> Institute of Physics, Universidade Federal de Mato Grosso do Sul, 79070-900 Campo Grande, MS, Brazil

<sup>3</sup> School of Computing, Engineering and Mathematics, Western Sydney University, Building XB, Room 2.48, Kingswood, Penrith Campus, Locked Bag 1797, Penrith, NSW 2751, Australia

**Keywords** Independent component regression · Principal component regression · Multiple linear regression · Water demand · Brazil

## 1 Introduction

Water is the vital resource to the living beings. Moreover, it is one of the most important elements in urban developments including water supply, stormwater, rainwater and wastewater systems. However, this invaluable resource is under stress at many locations around the globe due to combination of factors such as rapid urbanisation, industrial development, population growth and climate change. Consequently, water supply systems in many countries have become severely stressed to supply adequate water to the communities. These water stress conditions are likely to be exacerbated in already stressed countries and likely to be expanded to other countries under changing climatic conditions (Adamowski et al. 2010; Pingale et al. 2014; Proença de Oliveira et al. 2015; Haque et al. 2016). Therefore, an integrated, collaborative and adapting water resources management system is necessary to manage water demand effectively and sustainably. Forecasting water demand accurately is one of the means, among many others, to achieve more efficient and sustainable water resources management system. It can assist in achieving informed decisions to efficiently and effectively operate and manage water supply systems, as well as to prepare better planning and design of water supply systems in the long run (Bougadis et al. 2005).

Forecasting of water demand problem can be categorised into three types based on the forecast horizon (i.e. the duration of future period for what water demand is to be predicted) and periodicity (i.e. the time steps taken into the model), such as (i) short term forecasting, (ii) medium term forecasting and (iii) long term forecasting (Billings and Jones 2011; Donkor et al. 2012). There is no universal definition on the types of forecasting; however, some studies have defined that if the demand forecast exceeds the time frame more than two years it would be a ‘long term forecasting’, if the demand forecast lies between three months to two years it can be considered as a ‘medium term forecasting’, and it would be a ‘short term forecasting’ if the duration of forecasting period remains less than three months (Billings and Jones 2011). Long term forecasting is useful for developing policies and strategies to ensure adequate water supply in future. It assists in making decisions on the development, planning, and design of new water supply system infrastructure and in determining the efficient water conservation measures (Babel et al. 2007; Ghiassi et al. 2008; Firat et al. 2009; Herrera et al. 2010; Haque et al. 2014a). Medium term forecasting is valuable for taking strategic decision on the investment planning and the expansion of existing water infrastructures, while short term forecasting is necessary for effective operation and maintenance of water supply systems (Jain and Ormsbee 2002; Herrera et al. 2010). Therefore, it is apparent that all the water demand forecast horizons (i.e. 1 h to 10–20 years) are needed by water authorities to enable them to manage water supply systems effectively and efficiently.

Forecasting of accurate water demand is a challenging and difficult task. Several issues in combination make the demand forecasting challenging such as the nature and quality of available data, numerous water demand variables, diversity in forecasting horizons, geographical differences in the forecast areas and presence of various demographic conditions. These issues have motivated a number of studies to come up with better water demand modelling and forecasting tools in order to improve the overall forecast reliability. A variety of techniques have been adopted in water demand forecasting such as regression analysis (Hoffmann et al.

2006; Babel et al. 2007; Dziegielewski and Chowdhury 2011; Haque et al. 2014b), time-series modelling (Smith 1988; Zhou et al. 2000; Gato et al. 2007) and artificial neural networks (Diamantopoulou et al. 2005; Al-Zahrani and Abo-Monasar 2015; Perea et al. 2015; Mouatadid and Adamowski 2016). Some of the studies have also explored hybrid modelling by taking into account of two methods in combinations (Pulido-Calvo and Gutiérrez-Estrada 2009). Among these techniques, multiple linear regression is one of the most widely used techniques for water demand forecasting (Adamowski and Karapataki 2010) as the technique is comparatively simple and can be easily understood. Several forms of MLR techniques have been adopted such as linear, log-linear and log-log in water demand modelling. In those MLR models, the variables that are likely to influence water demand are taken into the model with or without log transformation.

Few studies have adopted the principles of MLR technique but used some modified water demand variables instead of original variables. For example, Haque et al. (2013) used principal component regression (PCR) technique to model and forecast water demand in the Blue Mountains area in Sydney, Australia. They adopted principal component analysis (PCA) technique to derive the principal components (PCs), which are the linear combination of the original variables, and thereafter incorporated those PCs into the MLR model to develop PCR model. They found that PCR model performed better than the MLR model in simulating the water demand. It should however be mentioned that PCR is not a new technique; it has been adopted in many water and environmental problems (e.g. Sousa et al. 2007; Rajab et al. 2013; Viswanath et al. 2015; Gulgundi and Shetty 2016). Of relevance, applications of PCR in water demand modelling is limited. For example Haque et al. (2013) and Koo et al. (2005) and Choi et al. (2010) have adopted PCR in the application of water demand forecasting among limited applications of PCR. Hence, it is important to explore the applicability of PCR in water demand forecasting, which will form the basis of a new tool in improving forecast reliability.

In PCA analysis, it is often needed to do rotation of the axes to identify the influential variables and to interpret underlying structures of the modelling variables. However, statistically independent structures are not always guaranteed in PCA due to the use of variance as objective function. As a further development of PCA, independent component analysis (ICA) has drawn attention to the researchers due to its potential to extract mutually independent components from explanatory variables (Comon 1994; Hyvärinen et al. 2004). It has been greatly developed in recent years as a potential statistical technique for blind source separation (Hyvärinen and Oja 2000; De Lathauwer et al. 2000). It can competently extract the independent components from the observed mixture of signals without any prior knowledge of the source signals by adopting the high order statistical characteristics of the source, that is, the fourth-order central moment. ICA has been widely adopted in signal processing arenas, such as image processing, financial analysis and biomedical signals (Vigário 1997; Hyvärinen 1999; Stone 2002; Makeig et al. 2002). Its application in speech recognition, telecommunication, spectroscopy and process monitoring has also been explored by several researchers (Westad and Kermit 2003; Yoo et al. 2004). Similar to PCR method (where PCs are incorporated into the MLR model), Independent Component Regression (ICR) method (where ICs are incorporated into the MLR model) were proposed by Chen and Wang (2001). Subsequently, ICR has been explored and adopted by several studies in various fields of engineering; for example, Westad (2005) applied ICR on sensory data, Kaneko et al. (2008) applied the technique to model aqueous solubility, and Lu et al. (2009) adopted ICR in financial forecasting. However, to the best of authors knowledge, application of ICR has not been explored in water demand forecasting.

Therefore, the present study is sought to explore, for the first time, the use of the ICR method for medium term urban water demand forecasting. It also compares the performance of developed ICR model with two other commonly adopted techniques, PCR and MLR model. The main innovation of this paper is in the adaptation of the powerful features of ICR method to water demand forecasting problems by extracting the independent components from the observed mixture of water demand related variables through the fourth-order central moment. It is expected that many future water resources forecasting studies will explore the applicability of ICR to enhance prediction accuracy of the models.

## 2 Study Area and Data

The study uses data from Aquidauana city in Brazil. Aquidauana is located in the south of the Midwest Brazilian region, in the Pantanal of South Mato Grosso (wetlands), which is a micro-region of Aquidauana. It is located at latitude  $20^{\circ}28'15''$  South and longitude  $55^{\circ}47'13''$  West, at an altitude of 149 m. It is situated between the Piraputanga and the Maracaju mountain ranges. Its territory is divided into two parts: the low one (two-thirds of the town) and the high one (in the mountain ranges).

The tropical climate of the region, with an annual average temperature of  $27^{\circ}\text{C}$ , features two opposing characteristics, namely the period between October and April is marked by floods and high temperatures, while from mid-July to end of September, is represented by a period of drought, with frosts and milder temperatures of approximately  $15^{\circ}\text{C}$ . It occupies an area of  $16\,958\text{ km}^2$ .

Monthly maximum temperature, relative humidity, wind speeds, rain, number of water consumers and water consumption data from January 2005 to 2014 were obtained from SANESUL System (Water Systems of South Mato Grosso). The meteorological data were obtained from the Water Resources Monitoring Center of South Mato Grosso – CEMTEC.

## 3 Methods

### 3.1 Multiple Linear Regression

Multiple linear regression attempts to model the relationship between two or more independent variables with a dependent variable by fitting a linear equation to the observed data. The general model form in MLR can be expressed as below:

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (1)$$

where  $Y$  is dependent variable,  $a_i (i = 0, \dots, n)$  are the regression coefficients generally estimated by least squares method and  $x_i (i = 0, \dots, n)$  are the independent variables.

The following assumptions are associated with MLR model (fitted by least squares method):

- (i) There is a linear relationship between dependent and independent variables;
- (ii) The error term  $\epsilon$  is a random variable that follows normal distribution;
- (iii) The mathematical expectation of error term,  $\epsilon$  is zero (i.e.  $E(\epsilon) = 0$ );

- (iv) Variance of error term is constant (i.e. homoscedastic assumption); and
- (v) There is no presence of high multicollinearity between any independent variables.

### 3.2 Principal Component Regression

Principal component analysis transforms the original data set of  $n$  variables, which are correlated to various degrees to a new data set containing  $n$  numbers of uncorrelated variables. These new variables are called principal components (PCs). The PCs are linear functions of the original variables in a way that the sums of the variances are equal for both the original and new variables. The PCs are sequenced from the highest variance to the lowest variance i.e. the first PC explains the highest proportion of variance in the data. The next highest variance is explained by the second PC and so on for all  $n$  PCs. The values of sPCs can be obtained by equations such as Eqs. 1 and 2. Although, the number of PCs and original variables are equal, normally most of the variance in the data set is explained by the first few PCs, which can be used to represent the original observations to a sufficient degree (Olsen et al. 2012). This helps in reducing the dimensionality of the original data set.

$$PC1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = \sum_{j=1}^n a_{1j}x_j \quad (2)$$

$$PC2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = \sum_{j=1}^n a_{2j}x_j \quad (3)$$

Where  $x_1, x_2, \dots, x_n$  are the original variables in the data set and  $a_{ij}$  are the eigenvectors.

The eigenvalues are the variances of the PCs and the coefficients  $a_{ij}$  are the eigenvectors extracted from the covariance or correlation matrix of the data set. The eigenvalues of the data matrix can be calculated by Eq. 4, as shown below:

$$|C - \lambda I| = 0 \quad (4)$$

Where  $C$  is the correlation/covariance matrix,  $\lambda$  is the eigenvalue and  $I$  is the identity matrix.

The PC coefficients or the weights of the variables in the PC are then calculated by Eq. 5:

$$|C - \lambda I| a_{ij} = 0 \quad (5)$$

In the PCR analysis, MLR and PCA are combined together to establish a relationship between the dependent variable and the selected PCs of the input variables (Pires et al. 2008). Mainly principal component scores obtained from the PCA are taken as the independent variable in the multiple linear regression equations to perform the PCR analysis. The general form of PCR model is as follows:

$$Y = \alpha + \beta_1 PC_1 + \beta_2 PC_2 + \dots + \beta_n PC_n \quad (6)$$

where  $Y$  is the dependent variable,  $\alpha$  is the model intercept,  $\beta$ 's are the regression coefficients and  $PC$ 's are the principal components.

### 3.3 Independent Component Regression

ICA is a statistical technique for decomposing observed multivariate data into statistically independent components expressed as the linear combinations of observed variables with minimum loss of information. The ICA bilinear model can be represented by the following equation (Parastar et al. 2012):

$$X = AS + E \quad (7)$$

where  $X$  is the observed data matrix,  $S$  and  $A$  represent the independent components and the coefficient matrix, respectively, this may be called as the mixing matrix of the ICs, and  $E$  is the error matrix.

The two main assumptions associated with ICA are as follows:

- (i) The independent components are statistically independent, and
- (ii) The independent components must have non-Gaussian distributions.

The objective of ICA is to identify a proper linear representation of non-Gaussian vectors in order to have the estimated vectors as independent as possible and to represent the mixed data as a linear combination of the independent components. The ICA model is quite similar to the PCA model where the multivariate data are represented by the linear combination of some orthogonal PCs. The difference is in the way of linear representations; ICA seeks to find ICs whereas PCA seeks to find orthogonal PCs. Independence is a much stronger condition than orthogonality because of its characteristics as high order statistics; therefore, ICA is generally considered to be more powerful than PCA in analysing multivariate data sets, as it can imitate the inherent properties of the original data sets in a better way (Hyvärinen et al. 2004).

ICR modelling is a combination of two statistical techniques, ICA and MLR. When ICA produces ICs from the original observed data sets then these ICs are incorporated into the MLR model as a replacement of the original variables to develop ICR model, which can be represented by the following equation:

$$Y = \alpha + \beta_1 IC_1 + \beta_2 IC_2 + \dots + \beta_n IC_n \quad (8)$$

where  $Y$  is the dependent variable,  $\alpha$  is the model intercept,  $\beta$ 's are the regression coefficients and  $IC$ 's are the independent components.

### 3.4 Performance Indices

The relative performance of the developed models were evaluated using four statistical criteria: the coefficient of determination ( $R^2$ ), root mean square error (RMSE), mean absolute relative error (MARE) and the Nash-Sutcliffe efficiency (NSE), as defined below:

- (i) Coefficient of determination ( $R^2$ ) measures the degree of collation between the observed and modelled values, and varies from 0 to 1. It indicates the strength of the model in developing a relationship among the dependent and independent variables. The

higher the  $R^2$  value, the better is the performance of the developed model.  $R^2$  can be calculated by the following equation:

$$R^2 = \left[ \frac{\sum_1^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_1^n (O_i - \bar{O})^2} \sqrt{\sum_1^n (P_i - \bar{P})^2}} \right]^2 \quad (9)$$

where,  $n$  is the number of observations,  $O_i$  and  $P_i$  are the observed and modelled water demand values at time  $i$ , respectively, and  $\bar{O}$  and  $\bar{P}$  are the mean of observed and modelled values, respectively.

- (ii) Root mean square error (RMSE) measures the variance of errors independently of the sample size and provides a good measure of model performance across the entire range of the data set. The smaller the value of RMSE, the better is the performance of the model with a perfect RMSE value of zero. RMSE is expressed by the following equation:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (10)$$

- (iii) The mean absolute relative error (MARE) indicates overall agreement between observed and modelled values. It considers all deviation from the observed values to the modelled values equally without considering the sign of the error (i.e. it takes absolute values into account). Therefore, it is always a position number and the smaller the MARE value, the better is the model performance. MARE value equal to zero indicate a perfect model. It can be expressed by the following equation:

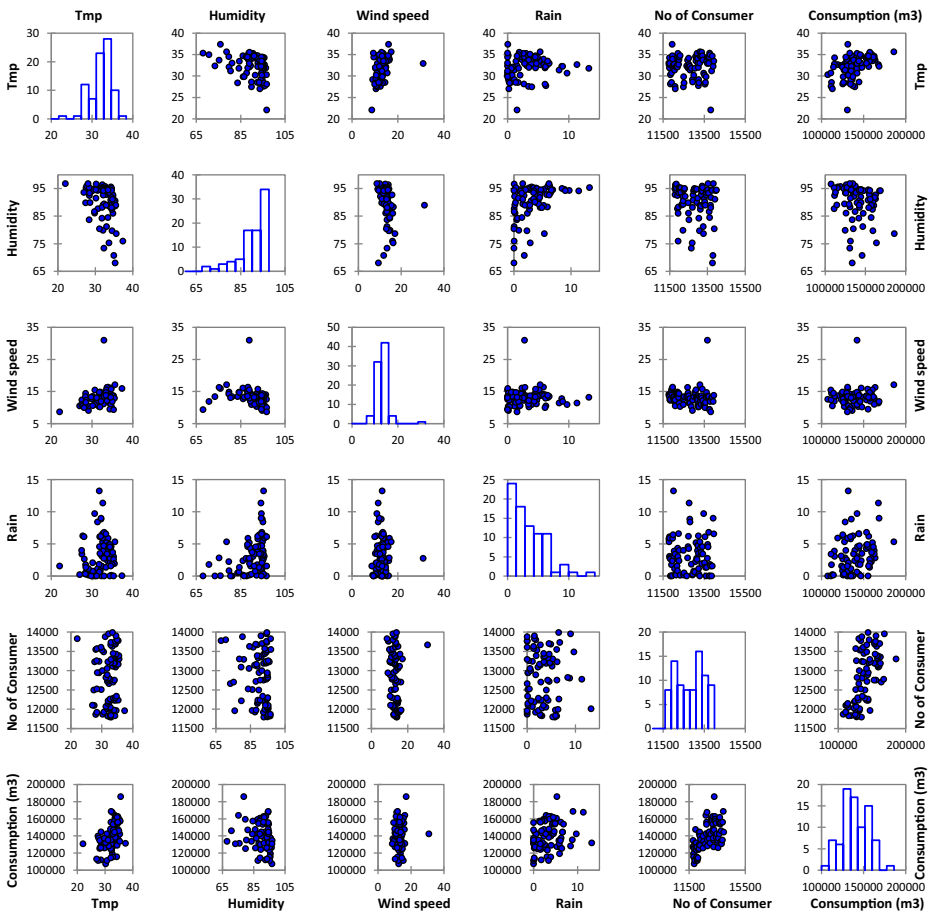
$$MARE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \quad (11)$$

- (iv) The Nash-Sutcliffe coefficient of efficiency is a normalized measure ( $-\infty$  to 1), that estimates the relative magnitude of the residual variance compared to the observed data variance (Nash and Sutcliffe 1970). An ideal value of NSE is one, which indicates a perfect model. A NSE value of zero indicates that the model results are as accurate as the mean of the observation. It can be calculated by the following equation:

$$NSE = 1 - \left[ \frac{\sum_1^n (O_i - P_i)^2}{\sum_1^n (O_i - \bar{O})^2} \right] \quad (12)$$

## 4 Results and Discussion

The correlation plot of the independent variables and dependent variables (i.e. water consumption) is presented in Fig. 1. It can be seen that water consumption is positively correlated with temperature and number of consumers, thus indicating that water consumption is higher if these two variables increase. Water consumption is found to be negatively correlated with humidity, indicating that if humidity increases water consumption would decrease. Water consumption shows no correlation with wind speed as shown in Fig. 1. It has somewhat positive correlation with rain, which means



**Fig. 1** Correlation matrix of the dependent and independent variables

that if rain increases, water consumption will also increase. But in practical situation, it should be opposite, as rain reduces the need for watering in the garden, and hence reduces water demand. From the rain histogram, it can be seen that rain amount is not significant in the study area. In addition, relation of water consumption with rain would be better captured in a daily time steps rather than monthly time steps. These two might be the reason for showing unusual correlation of rain with water consumptions, which is left for future research.

Correlations among the independent variables indicate that temperature is negatively and positively correlated with humidity and wind speed, respectively. It shows no significant correlation with rain. All the correlations are mentioned in this section are statistically significant at 10 % level. Humidity shows negative correlation with wind speed and positive correlation with rain. Wind speed shows no relation with rain. Furthermore, number of consumer is found to be uncorrelated to any of the climate variables (i.e. temperature, rain, wind speed and humidity), which is assumed to be reasonable. These correlations among the independent variables are found to follow the natural processes.

The developed water demand forecasting models by MLR, PCR, and ICR techniques are presented in Table 1. In those models, water consumption is modelled using independent



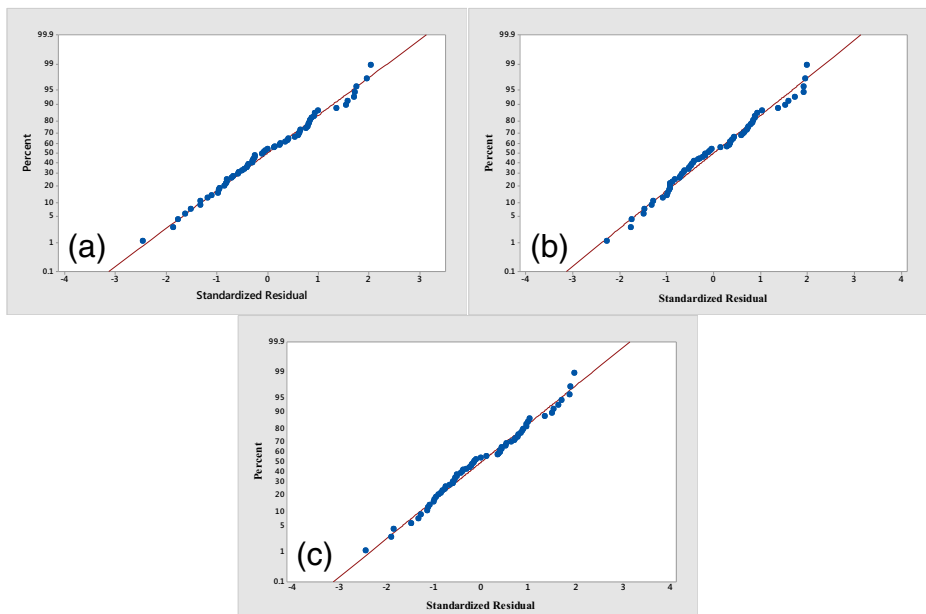
**Table 1** Developed water demand forecasting models

Models	Independent variables and coefficients	$R^2$ (%)	Standard error of estimate (%)	NSE	MARE (%)
Multiple Linear Regression (MLR)	$-143504 + 2286 * X_1 - 580 * X_2 + 1675 * X_3 + 20.3 * X_4$ where, $X_1$ = monthly maximum temperature, $X_2$ = monthly average humidity, $X_3$ = monthly total rainfall and $X_4$ = number of consumer	77.01	5.88	0.77	4.49
Principal Component Regression (PCR)	$143179 + 6038 * PC1 + 14644 * PC3$ where, $PC1$ and $PC3$ are the first and third principal components, respectively.	76.51	5.85	0.77	4.36
Independent Component Regression (ICR)	$141938 - 3811 * IC1 + 7017 * IC2 + 12748 * IC3$ where, $IC1$ , $IC2$ and $IC3$ are first, second and third independent components, respectively.	67.25	6.96	0.67	4.47

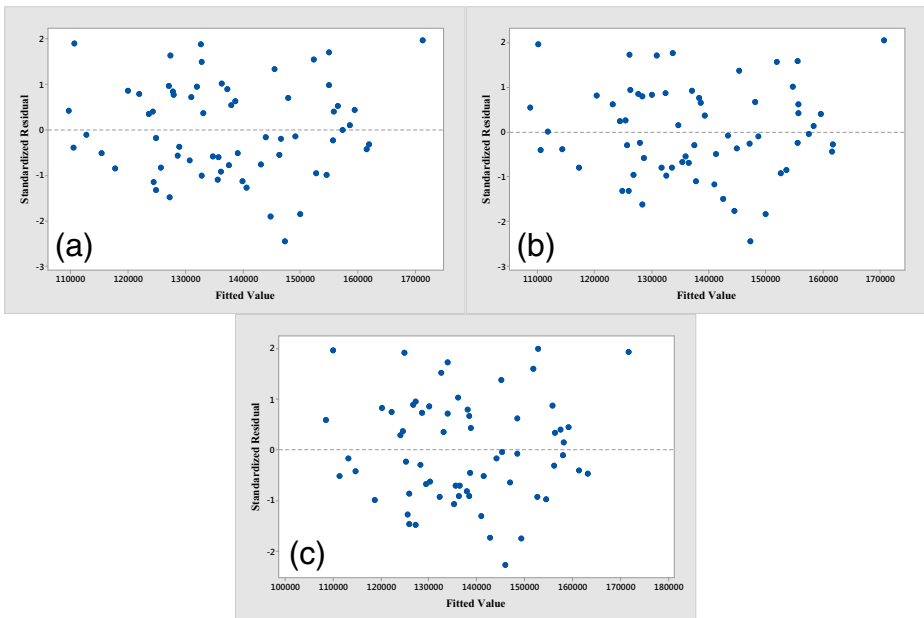
variables, and all the selected variables are statistically significant at 10 % significance level. It can be seen in Table 1 that model results are comparable to each other as the  $R^2$  and standard error of estimate values are similar for all of them. The calculated values of  $R^2$ , NSE and MARE of the developed models are found to be quite satisfactory, which indicate that water demand data has been fitted sufficiently by the models. The normal probability plot of the residuals (as shown in Fig. 2a–c) adopting the MLR, PCR and ICR show that most of the points are clustered around the blue line indicating that error terms are approximately normally distributed. In addition, the plots of fitted values vs. the standardised residuals for the MLR, PCR and ICR are presented in Fig. 3a–c, which show that for all the cases half of the data points are above the zero line and half of them are below zero line indicating that the error term has the zero mean value, which satisfy the regression assumption. Moreover, these plots indicate that the developed models satisfy the assumption of independence of error as no pattern for residuals has been detected.

The validation results of the developed water demand forecasting models using an independent data set are presented in Table 2. It can be seen that NSE values are negative for MLR and PCR models indicating that these two models have performed poorly in simulating the water demand for the independent period though the models have performed well in simulating the water demand for the model development period covering the data set. Of significance, the ICR model shows better accuracy than the MLR and PCR models as the NSE value is found to be 0.6, which can be deemed to be satisfactory. The MARE and RMSE results also indicate that ICR is better model than the other two.

The observed water demand vs. the simulated water demand values by the models are presented in Fig. 4, it can be seen that the MLR and PCR model results are similar and they have overestimated the demand for all of the months. On the other hand, ICR model perform



**Fig. 2** Normal probability of the residuals, a MLR, b PCR and c ICR



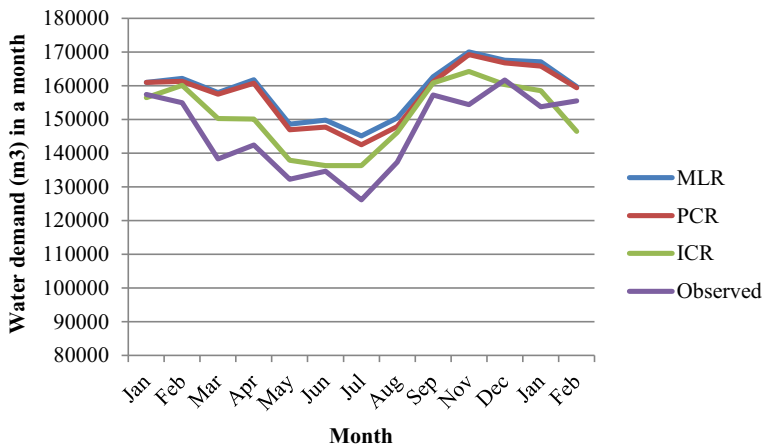
**Fig. 3** Plots of the fitted value vs. standardised residuals, **a** MLR, **b** PCR and **c** ICR

better than the MLR and PCR models as the simulated values are found to be close to the observed water demand values. However, the ICR model also shows overestimation bias in the results in most of the cases indicating that there is a room for improvement, which may be done by including more water demand variables (e.g. water price, income and evaporation) in the models. This issue has not been investigated in this study as this is beyond the scope of this study; rather the focus of this study is to compare the performance of the models based on the available water demand variables.

In a water demand modelling study, Haque et al. (2013) found that PCR model outperformed the MLR model in modelling water demand in the Blue Mountains region in Sydney, Australia. In this current study, it is found that PCR and MLR have performed in a similar manner but ICR has outperformed them. In the MLR, the water demand variables have taken as it is, without undertaking any transformation assuming that there is a linear relation existing between the water demand and the variables. In the PCR, all the variables are included in the principal components (PC), where PC1 accounts for the highest variance in data, then PC2 and so on. Generally the first few PCs are significant to explain most of the variance in the data. In the developed PCR model, PC1 and PC3 come as the significant variables among the 5 PCs. If PC2 is included in the model, it gives poorer results, and PC2 comes as statistically

**Table 2** Validation results of the developed water demand forecasting models

Models	NSE	MARE (%)	RMSE (m <sup>3</sup> )
Multiple Linear Regression (MLR)	-0.42	9.56	13464
Principal Component Regression (PCR)	-0.17	7.88	12226
Independent Component Regression (ICR)	0.60	4.26	7128



**Fig. 4** Observed water demand vs. modelled water demand by the MLR, PCR and ICR model for an independent period

insignificant ( $p$ -value comes as 0.5). Therefore, PC2 has not been included in the model. In the ICR, the generated independent components (IC) have been taken into the model. The IC's are independent to each other i.e. they generally have no relation among themselves, as by ICA the variables are transformed to equal number of separate components, which are independent to each other. Since ICs are free from collinearity and sovereign variables, the ICR model performs better than the other two models (MLR and PCR). However, the results are based on a limited quantity of data, which need to be extended with data from other cities in Brazil and around the world to make a better comparison of the ICR and PCR in water demand modelling and forecasting.

## 5 Conclusion

In this paper, a relatively new water demand forecasting modelling technique known as Independent Component Regression (ICR) technique is introduced. The ICR model is compared with two other commonly applied statistical models, Principal Component Regression (PCR) and Multiple Linear Regression (MLR) models. The modelling is done using data from the city of Aquidauana in Brazil. It has been found that the model results from the three techniques are generally comparable to each other. The validation results based on an independent data set indicate that the MLR and PCR models have performed poorly in simulating the water demand for the independent period. Interestingly, the ICR model is found to produce better accuracy than the MLR and PCR model as the NSE value is found to be 0.6, which can be viewed as satisfactory. The MARE and RMSE results also indicate that ICR is a relatively better model than the other two. Overall, it is concluded that ICR technique has the potential to develop successful water demand forecast models. However, some overestimation bias has been found in the results produced by the ICR model, which indicate that there is a room for further improvement, which may be achieved by incorporating additional water demand variables (e.g. water price, income and evaporation) into the models. The developed method can easily be adapted to other countries.

## References

- 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(10): 729–743
- Adamowski J, Adamowski K, Bougadis J (2010) Influence of trend on short duration design storms. *Water Resour Manag* 24(3):401–413
- Al-Zahrani M, Abo-Monasar A (2015) Urban residential water demand prediction based on artificial neural networks and time series models. *Water Resour Manag* 29(10):3651–3662
- Babel M, Gupta AD, Pradhan P (2007) A multivariate econometric approach for domestic water demand modeling: an application to Kathmandu, Nepal. *Water Resour Manag* 21(3):573–589
- Billings RB, Jones CV (2011) Forecasting urban water demand: American Water Works Association
- Bougadis J, Adamowski K, Diduch R (2005) Short-term municipal water demand forecasting. *Hydrol Process* 19(1):137–148
- Chen J, Wang XZ (2001) A new approach to near-infrared spectral data analysis using independent component analysis. *J Chem Inf Comput Sci* 41(4):992–1001. doi:10.1021/ci0004053
- Choi T-H, Kwon O-E, Koo J-Y (2010) Water demand forecasting by characteristics of city using principal component and cluster analyses. *Environ Eng Res* 15(3):135–140
- Comon P (1994) Independent component analysis, a new concept? *Signal Process* 36(3):287–314
- De Lathauwer L, De Moor B, Vandewalle J (2000) An introduction to independent component analysis. *J Chemom* 14(3):123–149. doi:10.1002/1099-128X(200005/06)14:3<123::AID-CEM589>3.0.CO;2-1
- Diamantopoulou MJ, Antonopoulos VZ, Papamichai DM (2005) The use of a neural network technique for the prediction of water quality parameters of Axios River in Northern Greece, 11/12, pp 52–62
- Donkor EA, Mazzuchi TA, Soyer R, Alan Roberson J (2012) Urban water demand forecasting: review of methods and models. *J Water Resour Plan Manag* 140(2):146–159
- Dziegielewski B, Chowdhury FJ (2011) Scenario-based forecast of regional water demands in Northeastern Illinois. *J Water Resour Plan Manag* 138(2):80–89
- Firat M, Turan ME, Yurdusev MA (2009) Comparative analysis of fuzzy inference systems for water consumption time series prediction. *J Hydrol* 374(3):235–241
- Gato S, Jayasuriya N, Roberts P (2007) Temperature and rainfall thresholds for base use urban water demand modelling. *J Hydrol* 337(3):364–376
- Ghiassi M, Zimbra DK, Saidane H (2008) Urban water demand forecasting with a dynamic artificial neural network model. *J Water Resour Plan Manag*
- Gulgundi M, Shetty A (2016) Identification and apportionment of pollution sources to groundwater quality. *Environ Process* 3(2):451–461
- Haque M, Rahman A, Hagare D, Kibria G (2013) Principal component regression analysis in water demand forecasting: an application to the Blue Mountains, NSW, Australia. *J Hydrol Environ Res* 1(1):49–59
- Haque MM, Rahman A, Hagare D, Kibria G (2014a) Probabilistic water demand forecasting using projected climatic data for Blue Mountains water supply system in Australia. *Water Resour Manag* 28(7):1959–1971
- Haque MM, Hagare D, Rahman A, Kibria G (2014b) Quantification of water savings due to drought restrictions in water demand forecasting models. *J Water Resour Plan Manag* 140(11):04014035
- Haque MM, Rahman A, Samali B (2016) Evaluation of climate change impacts on rainwater harvesting. *J Clean Prod* 137:60–69
- Herrera M, Torgo L, Izquierdo J, Pérez-García R (2010) Predictive models for forecasting hourly urban water demand. *J Hydrol* 387(1):141–150
- Hoffmann M, Worthington A, Higgs H (2006) Urban water demand with fixed volumetric charging in a large municipality: the case of Brisbane, Australia\*. *Aust J Agric Resour Econ* 50(3):347–359
- Hyvärinen A (1999) Sparse code shrinkage: denoising of nongaussian data by maximum likelihood estimation. *Neural Comput* 11(7):1739–1768
- Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. *Neural Netw* 13(4): 411–430
- Hyvärinen A, Karhunen J, Oja E (2004) Independent component analysis, vol. 46. Wiley
- Jain A, Ormsbee LE (2002) Short-term water demand forecast modeling techniques—Conventional methods versus AI. *J Am Water Works Assoc* 64–72
- Kaneko H, Arakawa M, Funatsu K (2008) Development of a new regression analysis method using independent component analysis. *J Chem Inf Model* 48(3):534–541. doi:10.1021/ci700245f
- Koo J, Yu M, Kim S, Shim M, Koizumi A (2005) Estimating regional water demand in Seoul, South Korea, using principal component and cluster analysis. *Water Sci Technol Water Supply* 5(1):1–7
- Lu C-J, Lee T-S, Chiu C-C (2009) Financial time series forecasting using independent component analysis and support vector regression. *Decis Support Syst* 47(2):115–125. doi:10.1016/j.dss.2009.02.001

- Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ (2002) Dynamic brain sources of visual evoked responses. *Science* 295(5555):690–694. doi:10.1126/science.1066168
- Mouatadid S, Adamowski J (2016) Using extreme learning machines for short-term urban water demand forecasting. *Urban Water J*. doi:10.1080/1573062X.2016.1236133, **published online**
- Nash J, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—A discussion of principles. *J Hydrol* 10(3):282–290
- Olsen RL, Chappell RW, Loftis JC (2012) Water quality sample collection, data treatment and results presentation for principal components analysis—literature review and Illinois River watershed case study. *Water Res* 46(9):3110–3122
- Parastar H, Jalali-Heravi M, Tauler R (2012) Is independent component analysis appropriate for multivariate resolution in analytical chemistry? *TrAC Trends Anal Chem* 31:134–143
- Perea R, Poyato E, Montesinos P, Díaz J (2015) Irrigation demand forecasting using artificial neuro-genetic networks. *Water Resour Manag* 29(15):5551–5567
- Pingale SM, Khare D, Jat MK, Adamowski J (2014) Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centers of the arid and semi-arid state of Rajasthan, India. *Atmos Res* 138: 73–90
- Pires J, Martins F, Sousa S, Alvim-Ferraz M, Pereira M (2008) Selection and validation of parameters in multiple linear and principal component regressions. *Environ Model Softw* 23(1):50–55
- Proença de Oliveira R, Saldanha Matos J, Monteiro AJ (2015) Managing the urban water cycle in a changing environment. *Water Util J* 9:3–12
- Pulido-Calvo I, Gutiérrez-Estrada JC (2009) Improved irrigation water demand forecasting using a soft-computing hybrid model. *Biosyst Eng* 102(2):202–218
- Rajab JM, MatJafri M, Lim H (2013) Combining multiple regression and principal component analysis for accurate predictions for column ozone in Peninsular Malaysia. *Atmos Environ* 71:36–43
- Smith JA (1988) A model of daily municipal water use for short-term forecasting. *Water Resour Res* 24(2):201–206
- Sousa S, Martins F, Alvim-Ferraz M, Pereira MC (2007) Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environ Model Softw* 22(1):97–103
- Stone JV (2002) Independent component analysis: an introduction. *Trends Cogn Sci* 6(2):59–64. doi:10.1016/S1364-6613(00)01813-1
- Vigário RN (1997) Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalogr Clin Neurophysiol* 103(3):395–404
- Viswanath N, Kumar P, Ammad K, Kumari E (2015) Ground water quality and multivariate statistical methods. *Environ Process* 2(2):347–360
- Westad F (2005) Independent component analysis and regression applied on sensory data. *J Chemom* 19(3):171–179
- Westad F, Kermit M (2003) Cross validation and uncertainty estimates in independent component analysis. *Anal Chim Acta* 490(1):341–354
- Yoo CK, Lee J-M, Vanrolleghem PA, Lee I-B (2004) On-line monitoring of batch processes using multiway independent component analysis. *Chemom Intell Lab Syst* 71(2):151–163. doi:10.1016/j.chemolab.2004.02.002
- Zhou SL, McMahon TA, Walton A, Lewis J (2000) Forecasting daily urban water demand: a case study of Melbourne. *J Hydrol* 236(3):153–164