Revista GEOGRÁFICA VENEZOLANA
Volumen 59(1)2018 enero-junio
ISSNe 2244-8853
RECIBIDO: diciembre, 2016 ACEPTADO: marzo, 2017
pp. 28—41

La demanda de agua

mediante la prevención de análisis de componentes principales en la ciudad de Aquidauana, Mato Grosso do Sul (MS), Brasil

Water demand through prevention of principal component analysis in the city of Aquidauana, Mato Grosso do Sul (MS), Brazil

Amaury de Souza[1]

Flavio Aristone[1]

Débora Aparecida da Silva Santos[2]

Wanderlei Mendes Ferreira[1]

1 Universidade Federal de Mato Grosso do Sul (UFMS) Instituto de Física/ Campo Grande (MS), Brasil
2 Universidade Federal de Mato Grosso, Instituto de Ciências Exatas e Naturais. Rondonópolis (MT), Brasil
amaury.de@uol.com.br; deboraassantos@hotmail.com

Resumen Conjuntos de datos colectivos de más de diez años (2005 a 2014) en Aquidauana, ciudad el estado de Mato Grosso del Sur, Brasil, se estudiaron en un intento de evaluar y determinar las contribuciones de fuentes que afectan al consumo de agua. Una técnica precisa de regresión lineal múltiple (MLR) se preparó como una herramienta avanzada para el consumo de agua, modelización y previsión. Además, se utilizó el análisis de componentes principales (PCA) para simplificar y comprender la compleja relación entre los parámetros de consumo de agua. Siete componentes principales fueron considerados responsables de la estructura de datos, denominada provisionalmente como: consumo de agua, número de consumidores, temperatura, humedad, precipitación y velocidad del viento. La estacionalidad explica el 94 % de la varianza total para todos los conjuntos de datos. Por lo tanto, el uso de PCA como entradas mejoró la predicción del modelo MLR mediante la reducción de su complejidad y la eliminación de la colinealidad de datos. El valor R^2 en este estudio es 0,93 y el modelo indica que 94 % de la variabilidad se explica por las siete variables independientes utilizadas en el modelo.
Palabras clave: regresión de componentes principales; demanda de agua; región urbana.

Abstract *Collective sets of data over ten years (2005 to 2014) in the city of Aquidauana, state of South Mato Grosso, Brazil, were studied in an attempt to assess and determine the contributions of sources affecting the water consumption. A precise technique of multiple linear regressions (MLR) was prepared as an advanced tool for water consumption, modeling and forecasting. Furthermore, principle component analysis (PCA) was used to simplify and understand the complex relation among Water consumption parameters. Seven principle components were considered responsible for the data structure, provisionally named as consumption of water, number of consumers, temperature, humidity, precipitation and wind speed. Seasonality explains 94 % of the total variance for all data sets. Therefore, the use of PCA as inputs improved the MLR model prediction by reducing their complexity and eliminating data collinearity. R^2 value in this study is 0.93 and the model indicates that 94 % variability is explained by the seven independent variables used in the model.*
Key words: *Principal component regression; water demand; urban region.*

1. Introduction

Historically, the most common forecasting methods in the sanitation sector are based on multiple regression and auto regression models. Recently, research advances in Artificial Intelligence (AI) began to emerge research on application of AI techniques in consumer forecasting models, especially artificial neural networks (ANN). Studies on specific consumption forecast of water supply system design is based on the number of water connections provided and population growth data available through predetermined calculation methodologies within the regression techniques, as discussed by Maurice *et al.* (1971), Perry (1981) and Balling & Gober (2007) have used models based on linear regression and auto regression models for forecasting consumption.

Balling & Gober (2007) studied the annual water consumption from 1980 to 2004 in Phoenix, Arizona, and they verified that the consumption was affected by climatic variables. The correlation among water use, annual average temperature, annual precipitation and values of the average annual Palmer's index for hydrological drought were +0.55, -0.69, and -0.52, during the studied period. The annual water consumption increases due to high temperatures, low precipitation and dry weather. Multivariate analysis using monthly weather data indicate that the annual water consumption is mostly controlled by dry weather, autumn temperature, and summer rainfalls. Model coefficients indicate that the conditions of temperature, precipitation and/or drought certainly affect the consumption of water, although the response of annual water consumption to weather changes was relatively low for an urban environment. Most of water residential consumption is used for external purposes like mechanical irrigation systems, due to the fact that Phoenix is an arid city. Thus the weather and water consumption are connected by a complex set of behavioral processes, which are crucial for designing programs aiming the most efficient use of water in urban areas.

Zhou *et al.* (2001) studied the rating average interval of the daily maximum water consumption for one, two, three and five consecutive days, at Melbourne (Australia). The series of daily consumption was obtained assuming the average consumption of 356 liters per person a day. The study involved three steps: calibrating a demand simulation model for daily water consumption to the months of higher consumption; rating water consumption for a time series; and, finally, calculating the recurrence average interval of extreme events.

To predict the daily water consumption, Zhou *et al.* (2000) created a model based on a set of equations that represent the effects of four factors: the trend, seasonality, correlation and climate autocorrelation. The basic water use was estimated by months of lower consumption. The long-term trend of the basic consumption, year to year, was represented by a polynomial function time dependent. The seasonal use was shaped by seasonality and length of components considering the six months of summer and winter separately. The developed model was tested using a cross validation procedure and a series of independent data during the summer period. The model of efficiency was $R^2 = 89.6\%$ and standard error was ± 8, considered acceptable.

In addition, considering the daily consumption forecast, Maidment *et al.* (1985) developed a short-term forecasting model based on time series analysis using the model of Box/Jenkins. The model is based on three propositions, as follows: **(1)** The total consumption, normally divided into: **a)** basic consumption, which is not susceptible to time consumption and the average consumption observed during the winter months; and **b)** seasonal consumption, which is susceptible to time and observed as the difference between the basic consumption and the total consumption during the other months of the year; **(2)** In the absence of rain, the seasonal consumption follows a standard feature for the year and depends on temperature conditions; **(3)** The occurrence of precipitation causes an immediate drop in seasonal consumption, which gradually decreases over time. The data used in the model was daily values obtained for Austin, Texas, in 1975-1981 period, corresponding to 97 % of the variation in municipal daily water consumption during the period. Forecasts of daily consumption were carried out for a two-week period.

Maidment & Miaou (1986) applied Maidment *et al.* (1985) methodology for daily consumption in nine cities, three in Florida, three in Pennsylvania, and three in Texas. The average coefficients R^2 of determination were 0.96 in Texas, 0.73 in Florida and 0.61 in Pennsylvania. They concluded that, as a proportion of average annual consumption, the average seasonal consumption for the three cities in each state was 23 % in Texas, 15 % in Florida and 5 % in Pennsylvania. The water consumption response compared to precipitation and air temperature was similar to cities within each state. Moreover, in response functions there was a small impact in relation to the size of the city. The water consumption response in relation to rain firstly depended on precipitation occurrence and secondly on quantity. They also observed that there is a nonlinear response of water consumption in relation to temperature changes.

The forecast water demand, in advance of 24 hours, can be performed by a mathematical model that combines previous demand data and other information, such as weather forecasts. Zhou *et al.* (2002) proposed a method for the hourly demand forecast of water using 24h interval records of water usage and climate information. The model involves two modules: daily and hourly. Daily module consists of a set of equations that represent the effects of seasonality, climate correlation, and autocorrelation. The hourly module is designed to separate the estimated daily consumption of hourly consumption. The models were calibrated using hourly and daily data for a period of six years, and validated with independent data for a period of seven months. On this last period, the forecast time model explained 66 % of the variation of the peak hourly consumption with a standard deviation of 162 liters / person / day.

Specific studies on the effect of climate variables were performed splitting the year into two seasons: winter and summer (Howe & Linaweaver, 1967; Carver & Boland 1980). Other studies were performed using a patterned seasonal regression function for each month (Morgan & Smolen, 1976; Yamauchi & Huang, 1977; Cassuto & Ryan, 1979). Many variables were used in the models to assess the impact of weather on water use. The temperature and rainfall are the two most common weather variables, and possibly the most effective (Morgan & Smolen, 1976); Weeks & Mcmahon, 1973).

Vogel & Bell (1997) developed a model for regional hydroclimatological basin in the northeastern United States, which relates annual runoff to climate and geomorphological characteristics of 166 stations. The regional hydroclimatological model for basin is then combined with the analytic relation among storage, reliability, elasticity and the yield of the water supply system. The sensitivity of various performance index of the supply system such as performance, reliability and system elasticity is derived due to weather, hydrological and storage conditions. The model results allowed determining the sensitivity of the supply system behavior to the climate change regime. The case studies in four basins of New York and Massachusetts supply system indicate that the annual approach of simple regional model can reproduce the approximate results of a hydroclimatological model much more monthly detailed. Gato *et al.* (2007) used data from consumer staples and correlated these data to weekday and climatic factors such as temperature and precipitation. The results revealed that the basic consumption depends on climatic factors, being affected by day of the week and the weekend.

In this study, water demand modelling is done thru multiple linear regression and PCR techniques to estimate the future demand. The aims are to evaluate the performance of developed multiple linear regression and PCR models by estimating and comparing several goodness of fit statistics. PCs are obtained by undertaking the PCA of seven water demand variables.

2. Studied area and data

For this study, the consumption region of Aquidauana city was chosen with average daily water consumption of 381 liters / day. Temperature is one of the factors that can influence water consumption (McDonald, 2012). For this reason, monthly temperature data (average, minimum and maximum), relative humidity (average, minimum, and maximum), wind speeds, precipitation, coefficient of seasonality, number of water consumers and water consumption from January 2005 to 2014 were obtained from SANESUL System (Water Systems of South Mato Grosso). The meteorological data were provided by the Water Resources Monitoring Center of South Mato Grosso (CEMTEC).

Aquiduana is located in the south of the Midwest Brazilian region, in the Pantanal of South Mato Grosso (wetlands), micro-region of Aquidauana. It is located at latitude 20º28'15" South and longitude 55°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 of 27 °C, features two opposing moments. The period between October and April is marked by floods and high temperatures. While from mid-July to end of September, it is represented by a period of drought, with frosts and milder temperatures of approximately 15 °C. It occupies an area of 16,958.496 km².

3. Materials and methods

In this study, multiple linear regression (MLR) equation and PCA were combined together to perform PCR analysis. This PCR model was adopted to predict the future water demand. Brief description of PCA, MLR and PCR are given in the following sections.

3.1 Principal component analysis

Principal component analysis transforms the original data set of n variables that are correlated among themselves to various degrees to a new data set containing n number of uncorrelated 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 one. The first PC explains the highest amount 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 all the PCs can be obtained by the same set as Equations 1 and 2. Although the number of PCs and original variables are equal, normally most of the variance in the data set can be explained by the first few PCs that can be used to represent the original observations, Abdul-Wahab *et al.* (2005) and Olsen *et al.* (2012). This helps in reducing the dimensionality of the original data set. These two equations are for PC 1 and PC 2:

$$a_{11}x_1 + a_{12}x_{2\ldots} + a_{1n}x_n = \sum_{i=1}^{n} a_{1j}x_j \qquad (1)$$

$$PC_2 = a_{21}x_1 + a_{22}x_2 \ldots\ldots + a_{2n}x_n = \sum_{j=1}^{n} a_{2j}x_j \qquad (2)$$

where x_1, x_2, ..., x_n are the original variables in the data set and a_{jj} are the eigenvectors.

The eigenvalues are the variances of the PCs and the coefficients a_{jj} are the eigenvectors extracted from the covariance or correlation matrix of the data set. The eigenvalues of the data matrix can be calculated by equation 3 as shown below:

$$|C - \lambda I| = 0 \qquad (3)$$

where C is the correlation/covariance matrix, λ the eigenvalue and I is the identity matrix. The PC coefficients or the weights of the variables in the PC are then calculated by equation 4:

$$|C - \lambda I| a_{jj} = 0 \qquad (4)$$

Due to differences in the units of the water demand variables used in this study, a correlation matrix of the variables was used to obtain eigenvalues and eigenvectors. The eigenvectors multiplied by the square root of the eigenvalues produce an n×n matrix of coefficients, which are called variable loadings. Importance of each original variable to a particular PC is represented by these loadings. Furthermore, the sum of the products of the variable loadings and the values of original variables produce a new set of data values, which are called component scores. These scores can be used in the multiple linear equations as new variables to predict the future water demand.

3.2 Multiple regression analysis

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 equation of a MLR model can be expressed as below Montogomery *et al.* (2001):

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots a_nx_n \qquad (5)$$

where y is the dependent variable, a_i (i = 0,1,..., n) are the parameters generally estimated by least squares method and x_i (i = 0,1,..., n) are the independent variables.

3.3 Principal component regression (PCR)

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 equation to perform the PCR analysis. The general equation of PCR model is as follows:

$$Y = a_1 \times PC_1 + a_1 \times PC_1 + \cdots + a_n \times PC_n \quad (6)$$

3.4 Performance indexes

The behavior of this model for both, development and validation steps, was evaluated calculating the following statistical parameters: correlation coefficient (R), mean bias error (MBE), mean absolute error (MAE), root mean squared error (RMSE) and index of agreement (d2), given by equations (7), (8), (9), (10) and (11) below, respectively:

$$R = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2-\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}} \quad (7)$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \quad (8)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(\hat{y}_i - y_i)| \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (10)$$

$$d_2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(|y_i-\hat{y}_i|+|y_i-\bar{y}_i|)^2} \quad (11)$$

R provides the variability measure of the data reproduced in the model. As this test does not give the accuracy of the model, other statistical parameters must be reported. MBE indicates if the observed values are over –or under– estimated. MAE and RMSE measure residual errors, which give a global idea of the difference between the observed and modelled values. The values of d2 compare the difference between the mean, the predicted and the observed values, indicating the degree of error free for the predictions (Gardner & Dorling, 2000; Chaloulakou *et al.*, 2003).

4. Results

Pearson correlation matrices of the water demand variables are presented in **table 1**. Statistically significant correlation coefficients ($p < 0.05$) are highlighted in bold. The linear relationship between two variables and the existence of the collinearity between the independent variables can be identified from these coefficients. Correlation is significant at the 0.05 level.

As it can be seen in **table 1**, per dwelling residential consumption were negatively correlated with humidity. This result was expected as the rainfall increase; the water requirement for consumption would be less. Hence, total water consumption would be reduced. Furthermore, it is obvious that with the enhancement of water conservation programs and restriction levels, the total water consumption goes down. It is been determined that water consumption is positively correlated with number of consumers, seasonality, temperature, speed and precipitation **(Table 1)**. As these variables are mostly related with temperature, water consumption will be more if the temperature is relatively high in a day. High correlation coefficients were found between the independent variables, such as seasonality and temperature, which demonstrate the existence of multicollinearity between the variables.

The PCA was done on the seven independent variables to explain per dwelling water consumption level in Water Supply systems. **Table 2** and **3** summarizes the results of the PCA on seven independent variables with the amount of variance explained by each PC. Moreover, these first five PCs explained around 94 % of the total variation of variables in PCA. These five PCs were selected for PCR. Contribution of a particular variable within a PC is normally judged by its variable loadings value. The higher the loading of a variable, the more contribution is reflected by that variable within a particular PC. The bold marked loads in **table 3** indicate the high existing correlation between the variables and corresponding PC.

The PCA was done on these six independent variables to explain per dwelling water consumption level. **Tables 2** and **3** summarize the results of the PCA on the seven

Table 1 Pearson correlation matrix for different variables

	Consumption	Number of consumers	Seasonality	Temperature	Humidity	Speed	Precipitation
Consumption	1.00	---	---	---	---	---	---
Number of consumers	0.59	1.00	---	---	---	---	---
Seasonality	0.72	0.01	1.00	---	---	---	---
Temperature	0.51	0.06	0.65	1.00	---	---	---
Humidity	-0.23	-0.17	-0.16	-0.35	1.00	---	---
Speed	0.13	-0.12	0.32	0.34	-0.28	1.00	---
Precipitation	0.24	-0.08	0.34	0.17	0.33	0.08	1.00

Table 2 Variance explained by the PCs

Eigenvalues %	% of total variance	Accumulated eigenvalue	Cumulative %
2.82	0.40	2.82	40.32%
1.48	0.21	4.30	61.42%
1.06	0.15	5.36	76.50%
0.71	0.10	6.06	86.62%
0.45	0.06	6.51	93.05%
0.39	0.06	6.90	98.63%
0.10	0.01	7.00	100.00%

Table 3 Component loadings - correlations between original variables and the PCs

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Consumption	**0.53**	0.01	**0.55**	0.49	-0.19	0.30	0.23
Temperature	-0.13	0.30	-0.10	0.15	**-0.68**	0.20	**-0.60**
Humidity	0.13	**0.81**	0.10	0.04	0.08	**-0.54**	0.11
Speed	-0.25	0.47	-0.18	-0.11	0.08	**0.70**	0.42
Precipitation	-0.22	-0.17	-0.27	0.23	**-0.58**	-0.30	**0.61**
Number of consumers	-0.33	0.01	-0.21	**0.82**	0.40	-0.01	-0.14
Seasonality	**-0.68**	-0.02	**0.73**	-0.06	-0.04	-0.06	0.02

independent variables with the amount of variance explained by each PC. It can be seen from **table 2** that the first four PCs had eigenvalues higher than 1. Moreover, these first five PCs explained around 94 % of the total variation of variables in PCA. These four PCs were selected for PCR. Contribution of a particular variable within a PC is normally judged by its variable loadings value. The higher the loading of a variable, the more contribution is reflected by that variable within a particular PC. The bold marked loads in **table 3** indicate the high existing correlation between the variables and corresponding PC.

All water demand variables were included in the five selected PCs. However, only certain variables showed high loadings within each PC, such as the first PC was heavily loaded on seasonality, and the second PC was heavily loaded with humidity. Component score coefficients (eigenvectors) and the values of the original variables were then multiplied to obtain PC score values. These score values were used as independent variables in the stepwise multiple linear regression analysis to determine the most significant PCs for water demand prediction. Data from November 2005 to December 2014 were used to develop the PCR model. Then the model was used to forecast water demand for the period of January 2010 to 2014.

In PCR analysis, PC1, PC2, PC3, PC4 and PC5 were found to have significant ($p < 0.05$) linear relationship with per dwelling water consumption (PDRC). The developed PCR model can be written as:

$$PCR = 103851 + 0.18 * PC1 + 97630 * PC2 - 284 * PC3 - 300 * PC4 + \\ +330 * PC5 + 497 * PC6$$

As can be seen in **table 2**, the five PCs (i.e. PC1, PC2 and PC5) could explain 93 % of the variation in water consumption. PC1 and PC2 were found to be the most significant independent variables in the regression analysis as the standardized beta coefficients values are the highest and the second highest for these two PCs, respectively. PC2 had positive impact on water consumption while PC5 had a negative impact **(Table 4)** as the sign of the regression coefficients were found to be positive and negative. This implies that if the value of PC2 increases, water consumption would be expected to increase and water

Table 4 Results of analysis of regressions the intercept values, number of customers, seasonality, temperature, humidity, wind speed and precipitation by MLR models and PCR

Predictor	Log PDRC				PDCR			
	Coef	SE Coef	T	P	Coef	SE Coef	T	P
Constant	5,51542	0,03756	146,85	0	103851	16519,00	6,29	0,00
Number	-0,00003	0,0000022	-13,29	0	0,18	0,40	0,50	0,62
Seasonality	0,25558	0,01174	21,78	0	97630,0	7295,00	13,38	0,00
Temperature	0,00045	0,0004007	1,13	0,27	-284,00	398,10	-0,71	0,48
Humidity	0,00006	0,0001459	0,42	0,68	-300,00	373,60	-0,80	0,42
Speed	-0,00016	0,0002287	-0,69	0,50	330,00	290,70	1,14	0,26
Precipitation	-0,00006	0,0003492	-0,19	0,85	497,00	371,20	1,34	0,18

consumption would decrease as the value of PC2 increases. Therefore, a total increase in significant variables of PC5, namely seasonality, temperature and precipitation would lead to an increase in the water consumption level. On the other hand, an increase in the significant variables of PC5 (humidity and speeds) would lead to a decrease in water consumption level, as expected. However, linear effects of these variables were partially incorporated in the model were also included in speed and humidity is negatively correlated with water consumption. As level of water restriction goes up, the level of water consumption would be expected to decrease.

The comparison of observed and predicted water consumption values by the PCR model is presented in figure 1. The forecasted monthly water consumption values were found to be close to the observed values. Average relative error values for all of the predicting months were found to be 0.22 %, which indicates that the model is capable of forecasting monthly water demand with a high level of accuracy. Three forms of multiple linear regression techniques, linear, semi-log and log-log were adopted to develop the multiple linear regression models. In the linear model, the relationship between the dependent variable, per dwelling water consumption and the independent variables (e.g. rainfall, temperature) were assumed linear. In semi-log model, only the dependent variable was in logarithmic form, whereas in the log-log models all the independent variables and dependent variable were entered in logarithmic form in the regression equation. After checking the model performances of these three models, it was found that the semi-log model performed best to model the water demand. Therefore, the semi-log model was taken as the final model to report in this study. The variables were retained in the regression equation for which the regression coefficients were significant at 5 % significance level. It was found that all variables were found to be significant in the equations. The resulting equation for the developed semi-log model is given below:

$$\log(PDRC) = 5{,}52 - 0{,}00003 * \text{numbers of consumers} + 0{,}26 * \text{seasonality} + 0{,}00045 * \text{temperature} + 0{,}00006 * \text{humidity} - 0{,}00016 * \text{speed} - 0{,}00006 * \text{precipitation}$$

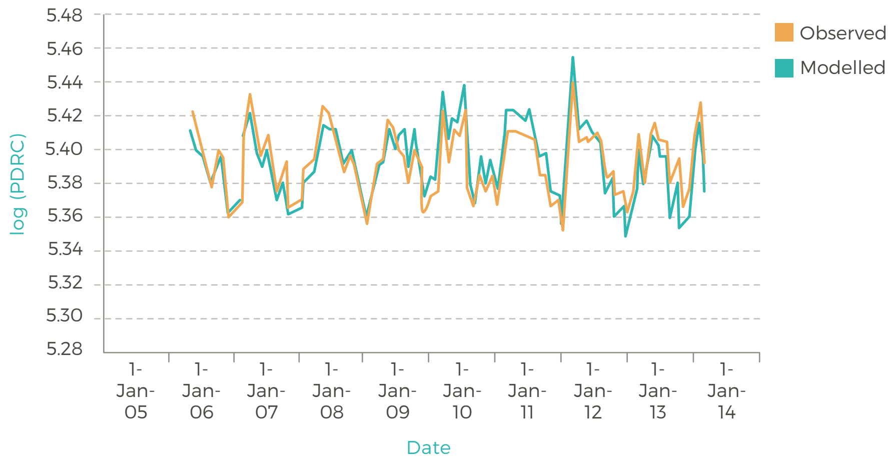

Figure 1 Comparison of predicted and observed water demand

The comparison of the performance of the developed MLR and the PCR model is presented in **table 5** for both the modelling and forecasting period. It was found that the performances of the models were nearly the same during the modelling period. However, the PCR model outperformed the MLR model during the forecasting period. All of the performance statistics were in favor of the PCR model. The PCR model considered the PCs as independent variables that accounted for the contribution of all the original variables without having any multicollinearity problem. On the other hand, in the developed MLR model, half of the original variables had to be discarded due to the multicollinearity problem, which might be the reason for the underperformance with respect to the PCR model.

Table 5 Results of the performance indices for both the MLR and PCR model

Performance Indices	MLR	PCR
R2	0.93	0,97
MBE	-0,001	-0,0007
MAE	0,02	0,019
RMSE	0,019	0,023
d2	0,98	0,99

5. Conclusion

In this study, the principal component regression (PCR) model was developed by combining multiple linear regression (MLR) and principal component analysis to identify the most important variables for water demand modelling and to forecast future water demand. It was found that all variables were the most significant independent variables in the PCR model. Therefore, the variables that had significant loadings within these PCs could be considered as important predictor variables for water demand forecasting. Therefore, the developed PCR model was used to forecast the future water demand, showing a high degree of prediction accuracy with an average relative error value 0.22 %. Moreover, the developed PCR model with seven PCs as independent variables was able to explain 93 % variation in water consumption level. The performances of the developed PCR model were compared to the MLR for both the modelling and forecasting period. Though both models performed similarly during the modelling period, the PCR model outperformed the MLR during the forecasting period.

6. References quoted

ABDUL-WAHAB, S. A.; BAKHEIT, C. S. & S. M. AL-ALAWI. 2005. "Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations". *Environmental Modelling & Software*, 20: 1.263-1.271.

BALLING JR., R. C. & P. GOBER. 2007. "Climate variability and residential water use in the city of Phoenix, Arizona". *Journal of Applied Meteorology and Climatology*, 46 (7): 1.130-1.137.

CARVER, P. H. & J. J. BOLAND. 1980. "Short and long run effects of price on municipal water use". Water Resources Research, 16(4): 609-616.

CASSUTO, A. E. & S. RYAN. 1979. "Effect of price on the residential demand for water within an agency". *Water Resources Bulletin*, 15(2): 345-53.

CHALOULAKOU, A.; SAISANA, M. & N. SPYRELLIS. 2003. "Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens". *Science of the Total Environment*, 313: 1-13.

GARDNER, M. W. & S. R. DORLING. 2000. "Statistical surface ozone models: an improved methodology to account for non-linear behavior". *Atmospheric Environment*, 34: 21-34.

GATO, S.; JAYASURIYA, N. & P. ROBERTS. 2007. "Temperature and rainfall thresholds for base use urban water demand modeling". *Journal of Hydrology*, 337: 364-376.

HOWE, C. W. & F. P. LINAWEAVER Jr. 1967. "The impact of price on residential water demand and its relation to system design and price structure". *Water Resources Research*, 3 (1): 13-32.

MAIDMENT, D. R.; MIAOU, S. P. & M. M. CRAWFORD. 1985. "Transfer function models of daily urban water use". *Water Resources Research*, 21 (4): 425-432.

MAIDMENT, D. R. & S. P. MIAOU. 1986. "Daily water use in nine cities". *Water Resources Research*, 22 (6): 845-851.

MAURICE, L. A.; SCOTT, T. & C. T. R. DONALD. 1971. *Treatise on Urban Water Systems*. Colorado State University. USA.

MCDONALD, C. J. 2012. *Factors that impact water intake of feedlot steers throughout the feedlot period*. Graduate Theses and Dissertations. Paper 12.400.

MONTOGOMERY, D. C.; PECK, E. A. & G. G. VINING. 2001. *Introduction to linear regression analysis*. Third edition, John Wiley & Sons, New York, USA.

MORGAN, W. D. & J. C. SMOLEN. 1976. "Climatic indicators in the estimation of municipal water demand". *Water Resources Bulletin*, 12 (3): 511-518.

OLSEN, R. L.; CHAPPELL, R. W. & J. C. LOFTIS. 2012. "Water quality sample collection, data treatment and results presentation for principal components analysis-literature review and Illinois River watershed case study". *Water Research*, 46 (9), 3.110-3.122.

PERRY, P. F. 1981. "Demand Forecast in Water Supply Networks. Journal of Hydraulic Division". *American Society of Civil Engineers*, 107: 1.077-1.087.

PIRES, J.; MARTINS, F.; SOUSA, S.; ALVIM-FERRAZ, M. & M. PEREIRA. 2008. "Selection and validation of parameters in multiple linear and principal component regressions". *Environmental Modelling & Software*, 23 (1): 50-55.

VOGEL, R. M. & C. J. BELL. 1997. "Fennessey, N.M. Climate, streamflow and water supply in the northeastern United States". *Journal of Hydrology*, 198: 42-68.

WEEKS, C. R. & T. A. MCMAHON.1973. "A comparison of water use, Australia and the US". *Journal of American Water Works Association*, 65 (4): 232-237.

YAMAUCHI, H. & W. HUANG.1977. « Alternative models for estimating the time series components of water consumption data". *Water Resources Bulletin*, 13 (3): 599-610.

ZHOU, S. L.; MCMAHON, T. A.; WALTON, A. & J. LEWIS. 2000. "Forecasting daily urban water demand: a case study of Melbourne". *Journal of Hydrology*, 236: 153- 164.

ZHOU, S. L.; MCMAHON, T. A.; WALTON, A. & J. LEWIS. 2002. "Forecasting operational demand for an urban water supply zone". *Journal of Hydrology*, 259: 189-202.

ZHOU, S. L.; MCMAHON, T. A. & Q. J. WANG. 2001. "Frequency analysis of water consumption for metropolitan area of Melbourne". *Journal of Hydrology*, 247: 72-84.

Lugar y fecha de finalización del artículo:
Campo Grande, Mato Grosso do Sul, Brasil;
noviembre, 2016.
Revisión: marzo de 2017