Analysis of domestic water demand variables of a residential colony in Ajmer, Rajasthan (India)
Ganpat Singh, Arun Goel and Mahender Choudhary

ABSTRACT
In this paper, significant variables of domestic urban water demand required for the purpose of estimation of urban water supply in five planned colonies of the City of Ajmer, Rajasthan, India, are identified. The data for these 16 variables are entered in the multiple linear regression (MLR) (stepwise) models in SPSS software, and domestic water demand models are developed. Based on these models, the six most significant variables, namely temperature (T), rainfall (RF), family size (FS), family income (FI), number of bathrooms (NB), and age of house (AH), are identified. The data of 16 variables are further utilized in principal component analysis (PCA), and five factors/variables are extracted, comprising combinations of these 16 variables. A regression coefficient of 0.76 is obtained in the PCA model. These six significant variables are further fed into a multilayer perceptron neural network (NN) model for water demand forecasting. The linear regression coefficient of NN is 0.90, very close to the MLR (stepwise) coefficient of 0.89, and verifying the dependence of water demand on these six variables. The outcome of the study suggests that the six variables are significant for estimation of water demand for Ajmer.

Key words | domestic water demand, multiple linear regression (MLR), neural network (NN), principal component analysis (PCA)

INTRODUCTION
Water demand forecasting is quite useful for the assessment of present water supply, upgradation of present resources, rate adjustment, capacity planning, future financial planning, scheduling of maintenance, and optimum operation of the water supply system (Billings & Jones 2008; Babel & Shinde 2011; Herrera et al. 2011; Choudhary et al. 2012; Odan & Reisl 2012). Problems like rapid and unplanned urbanization, environmental mismanagement, failed governance, demographic changes, and the scarcity of livelihood options for the poor are adding to the water scarcity problems in urban cities as mentioned by Field et al. (2012).

Many states in India have difficulty supplying enough clean water in cities. Rajasthan state, which lies in arid and semi-arid climatic zones, faces water supply problems in most of its urban areas. Major cities such as Jaipur, Bikaner, and Ajmer in Rajasthan depend on outside sources of water for domestic supply. The huge costs involved in inter-basin water transfers, along with conflicting water demands of various cities and user groups, make the sustainability of this approach questionable. Domestic water demand management offers a promising solution for sustainable urban water supply. To execute an effective demand management plan, it is necessary to identify a list of the parameters required to estimate the domestic water supply demand accurately in any water resources project.

LITERATURE SURVEY
Water demand management can be carried out in many ways, such as implementation of water demand reduction schemes, use of optimization techniques in supply systems,
development of sustainable water sources, as imposing effective water restrictions, as mentioned by Odan & Reisl (2022), Adamowski & Karapataki (2010), and Ghiassi et al. (2008). Forecasting urban water demand is crucial in managing water demand and supply, especially with the changes caused by climate change and population growth (Adamowski et al. 2013). The short-term water demand forecasts can be put to use in such projects as operational planning and management of an urban water distribution system, water conservation programme evaluation during drought conditions, and water pricing policy assessment, as mentioned by Odan & Reisl (2012). Billings & Jones (2008) stated that multilinear regression modelling can be applied for water demand forecasting which primarily depends on water price, ownership of water appliances, family income, family size, and urban density.

Khatri & Vairavamoorthy (2009) state that most demand forecasting studies have not considered uncertainties associated with future changes. They found that future water demand (at the monthly aggregate level) is primarily driven by change in population and by socio-economic changes, not climatic changes. They gave a detailed analysis on future water demand in Birmingham City and suggested appropriate planning options for maintaining a sustainable water supply. Yurdusev & Kumanhoglu (2007) focused on feasible water-saving measures vis-à-vis willingness at the domestic level; their results indicated that a 27% savings was manageable in the city of Manisa. Wu et al. (2010) stated that sustainable natural resources management might be achieved if taxes or charges were levied by volumetric consumption, thereby changing the behaviour of consumers towards volumetric use of water. Antonio & Mario (2011) found that climatic, sectoral, and technological modifications affect water consumption, but not income.

Various water demand forecasting methods and models have been developed in the past by many researchers; based on time series, autoregressive integrated moving average (ARIMA), regression methods, end-use method and decision support system, artificial neural networks (ANN), Institute of Water Resources IWR-Main (Water Demand Management Suite software), and fuzzy logic, etc. (Arbues et al. 2005; Billings & Jones 2008; Donkor et al. 2014; Singh et al. 2015). The outcome of these studies indicated that ANN, along with dynamic and hybrid models, perform better for short- to medium-term forecasting, while econometric models and the end-use method, coupled with simulation or scenario-based forecasting, may be used for long-term water demand forecasting (Jain et al. 2001; Donkor et al. 2014; Singh et al. 2015).

Choi et al. (2010) stated that water demand prediction based on factors analysis could be better when considering data for a large area including multiple cities instead of just one city. Ghimire et al. (2015) stated that the unavailability of individual data creates problems in assessing the impact of variability on household demand. However, their study was focused on price-related data, finding that the impact of an increase in price on water demand is very small during a drought period. Based on the brief literature review it can be observed that many researchers have worked on domestic water demand forecasting and other related studies at various places and timings. However, limited study has been undertaken in acute water scarce areas (arid zone) of India (i.e. Rajasthan) in the recent past related to domestic water demand forecasting in an urban city. Attempts are made in this study to assess the domestic water demand for urban areas of Ajmer City by collecting data on 16 variables for about 112 households. The collected data have been analysed using SPSS software based on principal component analysis (PCA), multiple linear regression (MLR) and ANN methods.

STUDY AREA

The historic city of Ajmer is situated in the centre of Rajasthan and lies about 135 km southwest of the state capital, Jaipur, as shown in Figure 1. Ajmer is settled in the cradle of the Aravali mountain range at an average elevation of 486 m above MSL. The city has a moderate climate with daily temperature varying from 26.9 °C to 46 °C during summer and 7.6 °C to 22.5 °C during winter. The average yearly rainfall is about 55 cm, and average humidity is 57%. The population of Ajmer is about 800,000 according to the 2011 census.

At present, Ajmer is dependent on Bisalpur Dam for its water supply, which is located about 120 km away from the city. Out of 108 ML per day (MLD) allocated for Ajmer, 87 MLD are required for domestic water supply, 0.812 MLD
for visitors, 9.58 MLD for institutional and commercial establishments, and 2.697 MLD for industry with a frequency of every other day.

**SURVEY DESIGN**

A questionnaire was prepared for the household survey and primary data were collected from Urban Improvement Trust (UIT) residential planned colonies of Ajmer City. Some secondary data were also collected from local body offices, such as the city’s Public Health Engineering Department, census data, and the Nagar Nigam office located in Ajmer City. The sample comprises 112 household responses collected from planned colonies of Ajmer, namely, Adarsh Nagar, Ajay Nagar, Panchsheel, HBU Nagar, and Vaishali Nagar. Samples are evenly distributed among seasons and the population. The variables considered in domestic water demand studies are mean monthly temperature (T) and rainfall (RF) occurrence or non-occurrence; socioeconomic and demographic variables are respondent age (RA), family size (FS), family income (FI), plot size (PS), house age (HA), water bill per year (WB), number of rooms (NR), number of bathrooms (NB), number of showers (NS), number of water closets (WC), number of washing machines (NW), number of AC units, number of coolers (NC), garden size (ft²), and domestic water demand (Qd) in kL per household per year. Data descriptions for all these variables are given in Table 1.

**RESEARCH METHODOLOGY**

The above primary data were collected via the random sampling method during 2014 and 2015. These data were

<table>
<thead>
<tr>
<th>Types of data</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
<th>Types of data</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly mean temperature (T)</td>
<td>14</td>
<td>36</td>
<td>26.95</td>
<td>Number of bathrooms (NB)</td>
<td>1</td>
<td>4</td>
<td>2.11</td>
</tr>
<tr>
<td>Rainfall (RF) (non-) occurrence</td>
<td>0</td>
<td>1</td>
<td>0.26</td>
<td>Number of WCs</td>
<td>0</td>
<td>6</td>
<td>1.97</td>
</tr>
<tr>
<td>Age of respondent (AR)</td>
<td>23</td>
<td>73</td>
<td>41.34</td>
<td>Number of showers (NS)</td>
<td>0</td>
<td>3</td>
<td>1.07</td>
</tr>
<tr>
<td>Family size (FS)</td>
<td>1</td>
<td>8</td>
<td>4.15</td>
<td>Size of garden (ft²)</td>
<td>0</td>
<td>450</td>
<td>56.12</td>
</tr>
<tr>
<td>Income per year (rupees) (FI)</td>
<td>132</td>
<td>2,220</td>
<td>764.67</td>
<td>No. of washing machines (NW)</td>
<td>0</td>
<td>1</td>
<td>0.82</td>
</tr>
<tr>
<td>Plot size (yd.²) (PS)</td>
<td>60</td>
<td>666</td>
<td>241.87</td>
<td>Number of coolers (NC)</td>
<td>0</td>
<td>6</td>
<td>1.63</td>
</tr>
<tr>
<td>House age (years) (HA)</td>
<td>2</td>
<td>50</td>
<td>11.02</td>
<td>Number of AC units</td>
<td>0</td>
<td>1</td>
<td>0.37</td>
</tr>
<tr>
<td>Water price per year (rupees) (WP)</td>
<td>156</td>
<td>1,800</td>
<td>331.28</td>
<td>Water demand (kL per capita per year) (Qd)</td>
<td>111.6</td>
<td>514.80</td>
<td>260.95</td>
</tr>
<tr>
<td>Number of rooms (NR)</td>
<td>1</td>
<td>5</td>
<td>2.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 | Location of study area.
extracted from the questionnaires and analysed through the methodological framework given in Figure 2. The primary aim of the present study is to develop a domestic water demand model and thence to identify a relationship between water demand and the most sensitive variables/drivers in the model. These relationships were further investigated using factors analysis (PCA), MLR (stepwise) methods, and a neural network (NN) (multilayer perceptron (MLP)) methods applied by using SPSS software. Before applying the above methodologies, the null hypothesis (normal distribution) was tested to determine whether it should be accepted or rejected.

CHECKING ADEQUACY OF DATA

Test of adequacy and sphericity

A combination of Kaiser-Meyer-Olkin (KMO) test and Bartlett’s test gives the sampling adequacy and sphericity. The KMO test measures the adequacy, which ranges from 0 to 1. The rule of thumb for the KMO test is that a value more than 0.5 is necessary before proceeding to further analysis (Hinton et al. 2012). The KMO value for our variables is 0.68, which shows that our sample is adequate. Bartlett’s test of sphericity gives the significance value as 0.000 ($p < 0.05$), indicating that our collected independent variables are qualifying for further regression analysis.

Test of normal distribution

The household data may be parametric or nonparametric; in case of doubt, all the data are considered nonparametric. The research hypothesis for collected data can be decided by the null hypothesis test. The data may be tested under a nonparametric test using the automatically combined one-sample hypothesis test. In this categorical variables are considered in a binomial test, and other continuous variables are taken into account for the Kolmogorov-Smirnov test (K-S test). The K-S test is tested with normal distribution significance levels. These hypotheses were tested at 90% confidence intervals (CI), and finally the null hypothesis was rejected for all considered variables.

SELECTION OF SIGNIFICANT VARIABLES

PCA (factors analysis)

In PCA, all the independent variables in a factor attempt to explain the maximum contribution of variables through their percentage variance. Cumulative percentage of variation is considered up to an eigenvalue greater than one. In this study, five components have an eigenvalue greater than 1.0, indicating that they have a significant contribution to estimation of domestic water demand.

The rotated components matrix has shown the dependence of these five components on 16 variables/factors in the analysis. Based on these components/factors, the goodness of fit of MLR coefficient $R^2$ of this model is 0.76, which is quite satisfactory for domestic water demand forecasting. Stepwise MLR models are further applied in this study because the most significant variables are identified and selected at every step of the regression process, showing the strongest relationships with domestic water demand.
MLR analysis

A MLR (stepwise) model was developed in the present study. The regression coefficient is decided by the significance of independent variables included in the model. However, this type of regression model is developed on multiple correlation coefficients (Pearson’s coefficients and their significance values), analysis of variance, probability, etc. The multiple correlation coefficients give the relationships between the dependent and predictor variables. Sixteen variables in MLR and output models were generated by stepwise consideration of significant variables otherwise excluded from the process. Most researchers have used regression methods as predictor forms from the relation of water demand and factors or drivers that affect it. The output of multiple regressions obtained by using SPSS software is discussed below.

In stepwise MLR, all 16 explanatory or independent variables are entered for analysis. This regression model is summarized in Table 2, in which six models were generated based on the importance of independent (predictor) variables. Model 1 includes the family size (number of persons) predictor and has the lowest $R = 0.82$, $R^2 = 0.67$, and SE = 43.64. This shows that water demand is highly dependent on family size. Model M2 includes family size and monthly mean temperature thereby increasing $R = 0.89$, and $R^2 = 0.79$ while decreasing SE = 35.03. Successively the model M6 includes most predictors – family size, monthly mean temperature, income per year (rupees), rainfall occurrence or non-occurrence, number of bathrooms, and house age (years) – and has the highest $R = 0.95$, and $R^2 = 0.89$ while SE = 25.66. Hence, Model 6 accounts for 89.30% of the variance of predictors/independents in the water demand.

In each successive step an increase in regression coefficient demonstrates a better model fit and thence reduction of the standard error. The software stops further checking of parameters when there is no significant increase in model fit or reduction in standard error.

Analysis of variance

In the analysis of variance (ANOVA) test, the F-statistic gives the statistical significance of the overall equation. The ANOVA tested the importance of regression Models 1–6; the F-statistic satisfied the significance values ($p < 0.005$), verifying that independent/predictor variables such as family size, monthly mean temperature, income per year (rupees), rainfall occurrence or non-occurrence, number of bathrooms, and house age (years) explain a significant amount of variance in water demand.

Analysis of regression coefficients

As discussed in the model summary, six regression models have been generated in this study. In the stepwise regression method, the regression coefficient gives the model-wise unstandardized coefficient (additive constant, predictors multiplying constant and standard error) and tests statistical significance values. According to Table 3, the highest regression coefficient ($R^2 = 0.89$) of Model M6 gives the domestic water demand ($Q_d$) in kL per household per year as shown in Equation (1):

\[
Q_d = 40.59(FS) + 4.77(T) + 0.04(FI) - 28.51(RF) + 18.73(NB) - 0.74(HA) - 91.84
\]  

(1)

Table 2 | Model summary (obtained by using SPSS software)

<table>
<thead>
<tr>
<th>Model no.</th>
<th>Variables included</th>
<th>$R$ (Pearson coefficient)</th>
<th>$R^2$ (Regression coefficient)</th>
<th>SE (Standard error of the estimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>FS</td>
<td>0.82$^a$</td>
<td>0.67</td>
<td>43.64</td>
</tr>
<tr>
<td>M2</td>
<td>FS &amp; T</td>
<td>0.89$^b$</td>
<td>0.79</td>
<td>35.03</td>
</tr>
<tr>
<td>M3</td>
<td>FS, T, &amp; FI</td>
<td>0.92$^c$</td>
<td>0.84</td>
<td>30.10</td>
</tr>
<tr>
<td>M4</td>
<td>FS, T, FI, &amp; RF</td>
<td>0.93$^d$</td>
<td>0.86</td>
<td>28.45</td>
</tr>
<tr>
<td>M5</td>
<td>FS, T, FI, RF, &amp; NB</td>
<td>0.94$^e$</td>
<td>0.88</td>
<td>26.04</td>
</tr>
<tr>
<td>M6</td>
<td>FS, T, FI, RF, NB, &amp; HA</td>
<td>0.95$^f$</td>
<td>0.89</td>
<td>25.66</td>
</tr>
</tbody>
</table>

$^a$(Constant), FS; $^b$(Constant), FS & T; $^c$(Constant), FS, T & FI; $^d$(Constant), FS, T, FI & RF; $^e$(Constant), FS, T, FI, RF & NB; $^f$(Constant), FS, T, FI, RF, NB & HA.
where \( T \) is in °C, \( FI \) in thousand rupees, \( RF \) is 1 if rainfall occurred at the selected occasion or 0 if not, and \( HA \) is in years. As shown in Table 2, the parameter family size is the most significant and is given a coefficient of 40.59 in the Equation (1). The second most important parameter of Table 2, mean temperature, is given a coefficient of 4.77. The sign of this multiplying coefficient depends on the nature of effect of that parameter on water demand. A positive sign shows that water demand increases with the increase in the value of parameter and vice versa.

**DISCUSSION OF RESULTS AND MANAGEMENT STRATEGIES**

The correlation output of the regression method represents the interrelationship between all independent variables among themselves and with domestic water demand. An additive constant of equation \(-91.84\) is obtained for the regression model. As per the correlated result, family size shows the highest Pearson’s coefficient, 0.82 (Table 2), indicating that water demand highly depends on the number of persons in a family. Various government schemes for birth control, family planning, and social awareness programmes can be effectively implemented to control family size.

The second-strongest correlation with water demand belongs to the daily mean temperature of the region; here the correlation coefficient is 0.89 (Table 2) and multiplying coefficient \( B = 4.77 \) (Table 3). Demand is greater in the summer, less in the winter. The effect of temperature is slightly overcome by providing green building concepts, such as having a green area surrounding a development and using recycled water outdoors. Solar passive architecture can control indoor temperature during the summer, thereby reducing water demand. Various adaptation/mitigation plans for climate change, such as creating bodies of water and increasing green cover in urban areas for heat absorption, can reduce household water demand.

The third-most significant variable from Table 2 affecting water demand is family income with \( B = 0.04 \) (Table 3): the higher a family’s income, the greater its demand for water (Ghimire et al. 2018), because higher-income families having more water-using appliances in their houses, such as washing machines, water purifiers, and reverse osmosis plants, as well as more faucets, bathtubs, etc. However, demand sustainability may be increased if high-income families purchase modern, high star rating (water-saving) appliances, such as dual flushing cisterns, automatic washing machines, and sensor-based water faucets, whether on their own or as a result of government nudging.

The fourth-most important parameter is the occurrence of rainfall from Table 2 and has a multiplying coefficient \( B = -28.51 \) (Table 3). Its relation is given negatively, i.e. demand is less when rainfall occurs, because on a rainy day, the temperature falls and humidity rises, so persons may take fewer baths or use the cooler less. To improve sustainability, the municipal body should divert the water supply in a controlled manner to the supply areas/colonies on particularly rainy days.

The fifth-most significant parameter is the number of bathrooms in the household: water demand increases with

**Table 3 | Calculated regression model with coefficients**

<table>
<thead>
<tr>
<th>Model 6</th>
<th>Unstandardized coefficients</th>
<th>Standardized coefficients</th>
<th>95.0% Confidence interval for B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. error</td>
<td>Beta</td>
</tr>
<tr>
<td>(Constant)</td>
<td>-91.84</td>
<td>17.03</td>
<td>-</td>
</tr>
<tr>
<td>Family size</td>
<td>40.59</td>
<td>2.28</td>
<td>0.72</td>
</tr>
<tr>
<td>Monthly mean temperature</td>
<td>4.77</td>
<td>0.45</td>
<td>0.37</td>
</tr>
<tr>
<td>Income per year (Rs)</td>
<td>0.04</td>
<td>0.01</td>
<td>0.18</td>
</tr>
<tr>
<td>Rainfall (non-)occurrence</td>
<td>-28.51</td>
<td>5.87</td>
<td>-0.16</td>
</tr>
<tr>
<td>Number of bathrooms</td>
<td>18.73</td>
<td>4.26</td>
<td>0.18</td>
</tr>
<tr>
<td>House age (years)</td>
<td>-0.74</td>
<td>0.36</td>
<td>-0.07</td>
</tr>
</tbody>
</table>
the number of bathrooms. Since the number of bathrooms is directly related to the number of water-using appliances in a house and to the total number of rooms in a house and the size of its plot, demand can be managed by constructing houses in equal sizes according to accommodation per person or family.

This regression result obtained for the current model is $R^2 = 0.89$ with $SE = 25.66$. This shows that the model is capable of representing 89% of the variation in the domestic water demand for the study area on the basis of six significant parameters with family size being the most significant of them all. Some of the past researchers have reported better results in their studies but that can be attributed to a more homogeneous sample size. In the current study the most significant variable of family size has a range of 1 to 8 (Table 1) and the heterogeneity of the sample size reduces the goodness of fit. The next most significant parameter, monthly mean temperature, also shows wide variation (Table 1) due to the geographical location of the study area in hot, arid climate. Arbues et al. (2005) concluded that the water bill is small relative to overall family income. In these urban colonies, the Public Health Engineering Department of Ajmer provides water at a subsidized rate. However, in some colonies, people have to pay an exorbitant price (up to Rs. 1,800 per month; Table 1) to water vendors while coping with shortages.

REGRESSION (STEPWISE) MODEL VALIDATION WITH ANN

The data collected were also analysed using the MLP NN modelling, in which the sigmoid function scale data between 0 and 1. As per the processing summary, in the ANN modelling 77 data points (68.8% of 112 total) were used for training and 35 (31.3%) for testing in the ANN modelling. For the training data, the relative error is 0.09 and the sum of square error is 3.75, whereas for the testing data, the relative error is 0.11 and the sum of square error is 1.14 in the ANN modelling. The linear $R^2 = 0.90$, is obtained from MLP (NN) and $R^2 = 0.89$ from the regression (stepwise) model. Figure 3 shows the model-wise comparisons of actual domestic water demand versus predicted domestic water demand for MLR (stepwise), MLR (PCA), and MLP NN, respectively.

Figure 3 | Comparison of actual domestic water demand and predicted domestic water demand.
CONCLUSION

Domestic water demand estimation is essential for proper planning and management of any water resources project. Accurate domestic water demand forecasting depends on a number of factors related to climatic, socio-economic, and demographic factors. The data for 16 variables affecting domestic water demand were collected from five urban colonies (112 households) of Ajmer City. By using statistics techniques in SPSS software, various checks of the adequacy of the data set and a check for normal distribution are applied. MLR with PCA is carried out on all 16 variables giving $R^2 = 0.76$. Out of 16 variables, 6 significant variables are identified and studied using MLR Stepwise. MLR Stepwise on six significant variables gives $R^2 = 0.89$ and SE = 25.66 using SPSS software. These results of MLR Stepwise are comparable with the earlier studies and an equation is proposed for estimation of domestic water demand can be applied successfully for Ajmer City. Some management strategies have also been proposed to tackle the problem of scarcity of water.

REFERENCES


