Water Practice & Technology Vol 13 No 3
doi: 10.2166/wpt.2018.082

Application of multiple regression analysis in projecting the water demand for the City of Cape Town

M. Lawens[a,*] and C. Mutsvangwa[b,†]

[a] Department of Civil Engineering, University of Applied Sciences Mainz, Mainz, Germany

[b] Department of Civil Engineering & Surveying, Cape Peninsula University of Technology, Cape Town, South Africa

*Corresponding author. E-mail: mirjam.lawens@hs-mainz.de

Abstract

Historically, the city of Cape Town has been affected by water shortages and it can be assumed that the situation will be exacerbated in the coming decades by a growing population, economic development and climatic changes as additional stress factors. In order to defuse the situation, the city of Cape Town has commissioned various feasibility studies concerning the implementation of alternative water sources, with as yet unpublished conclusions. Since sustainable water resource planning requires a comprehensive understanding of the water demand, the objective of this study was to predict the future demand by the city of Cape Town by analysing its significant drivers. For this purpose, a linear multiple regression analysis was applied on parameters which influence water demand, namely: population, economy, water losses and water restrictions. In order to establish the linear multiple regression model and its regression coefficients, historical data was used for the period 2001 to 2012. The result of the regression analysis showed that the water demand of the city of Cape Town is only decisively influenced by population and water losses. In addition, the model indicated that a new source would be required by 2021. Thus, water conservation and water supply strategies can be adapted accordingly to ultimately enable a sustainable management of the water sources in the city of Cape Town.

Key words: city of Cape Town, multiple regression analysis, water deficit, water demand

INTRODUCTION

Towards the end of the 1990s the city of Cape Town implemented an integrated water resource planning approach to deal with a threatening scarcity of water. As a result, various water conservation and water demand initiatives were implemented. In spite of this, the city has been affected by water shortages and it can be assumed that the situation will be exacerbated in the coming decades by a growing population, economy and climatic changes (Frame & Killick 2005). In order to defuse the situation, the city of Cape Town has commissioned various feasibility studies concerning the implementation of alternative water sources, with as yet unpublished conclusions (CoCT 2013a).

The Western Cape Province is situated along the Atlantic coast and hence has inadequate fresh water sources. The city of Cape Town is the provincial capital, with an estimated population of 3.7 million. The city has a Mediterranean type of climate with average rainfall ranging between 560 to 1,400 mm. The city's average winter and summer temperatures are 10 °C and 27 °C respectively (Tadross & Johnston 2012). The Department of Water Affairs has identified the Western Cape Province as the first region in South Africa likely to run out of fresh surface water resources (Otien &

† Formerly.

Water Practice & Technology Vol 13 No 3
doi: 10.2166/wpt.2018.082

Ogutu 2005). Currently groundwater and surface waters account for about 2% and 98% respectively, with network distribution losses of about 14.3%. According to the city of Cape Town water resources report (2013b), the current water demand is 27% below unconstrained projected water demand, attributed to water restrictions. Frame & Killick (2005) highlighted that the implementation of water demand management initiatives significantly reduced the unconstrained water use of the city of Cape Town. This notion is also supported by the Integrated Water Resource Planning Study commissioned by the city of Cape Town in order to respond to the constantly increasing demand for water due to population growth and poor rainfall in the region. Water demand management started in 2001 and was fully implemented in 2012. In the following years, the strategy was annually monitored and reviewed.

Water resource augmentation schemes took place with the opening of the Berg River Dam in 2009, which increased the water supply of the city of Cape Town. Indeed the opening of Berg River Dam increased the water supply of the city, but the need for further water resource augmentation schemes was only postponed. Other studies concerning the feasibility of implementing new water sources were commissioned (CoCT 2013a). However, at this juncture it is not clear whether or when these water supply schemes should be in place to meet the projected water demand.

Since sustainable water resource planning requires a comprehensive understanding of the water demand, the objective of this study was to predict the future demand by the city of Cape Town by analysing its significant drivers. For this purpose, a statistical analysis method was applied, namely a multiple linear regression analysis.

METHODS

The research method involved the collection of historical data for the period 2002 to 2012 and their statistical analysis. The data was provided by the city of Cape Town and contained detailed historical information on water demand since 2002, as well as the current water resources. Based on this data, the most significant influences on water demand were identified, namely: demography, economic growth, water losses and restrictions. Figure 1 shows the development of the population, economy, imposed water restrictions and water losses of the city of Cape Town for the period 2002 to 2012. Based on these influential variables on water demand, a linear multiple regression model was applied.

A multiple linear regression analysis is a statistical tool that examines the relationship between a dependent variable (criterion variable) and one or more independent variables (predictor variables). It relates to the question as to whether any relationship exists and if so, how good is the relationship. A general multiple regression model is given as (Reinboth 2006):

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n \quad (1)$$

where y is the dependent or target variable of the nth observation, x is the independent variable of the jth observation. The independent variable (x) will fluctuate with y in its unique way. The regression coefficients (b_n) of independent variable are respective changes in y per unit change and are unknown parameters. The regression constant (b_0) is the y intercept when $x = 0$. Taking into consideration the stated four influences on the water demand, the following specific regression model for the city of Cape Town which relates water demand with its influences was developed:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 \quad (2)$$

where y is water demand of the city of Cape Town depending on the variables (m^3/year), x_1 is the population of the city of Cape Town (capita), x_2 describes the water losses (m^3/year), x_3 is an economic indicator of the city of Cape Town as GDP per capita (ZAR/capita), x_4 describes the amount

Water Practice & Technology Vol 13 No 3
doi: 10.2166/wpt.2018.082

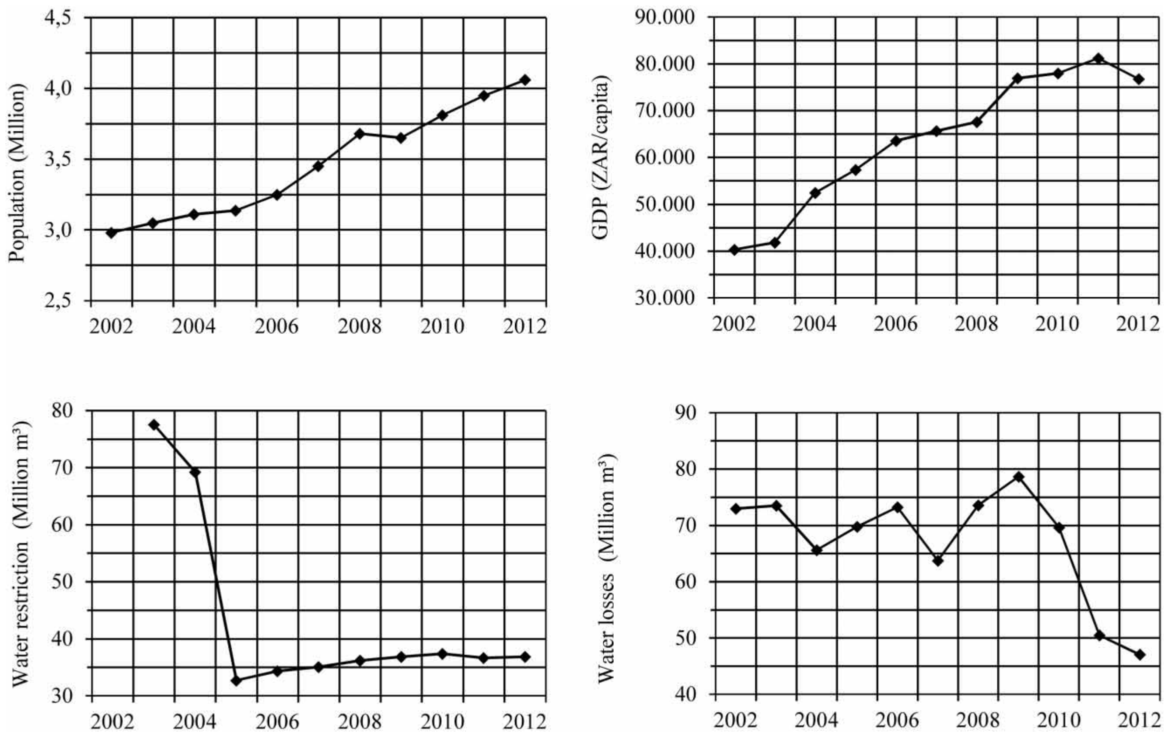

Figure 1 | **top left**: Population 2002–2012 (CoCT 2011) **top right**: Economy 2002–2012 (OECD 2008; CoCT 2013d) **bottom left**: Water restrictions 2002–2012 (CoCT 2008) **bottom right**: Water losses 2002–2012 (CoCT 2013d).

of water restrictions (m^3/year), b_0 is the regression constant and b_1 to b_4 are regression coefficients for x_1 to x_4 respectively. The regression analysis was conducted for the period 2002 to 2012, which coincided with the start of the implementation of the water demand strategies.

To identify the most suitable subset of predictors the backward elimination method was applied. For this purpose, a *p*-value test was conducted for each regression factor as part of the multiple regression analysis. The *p*-value indicates the possibility of mistakenly rejecting a null hypothesis. It can be between 0 and 1, and is expressed as a percentage. The significance level is described by α_{crit} and is normally 5%. The predictor with the highest *p*-value, which is greater than α_{crit}, is removed and the regression analysis is repeated (Farway 2002). This in turn means that if the *p*-value is less than 5%, the considered factor is assumed to be significant. The procedure is iterated until all *p*-values are smaller than α_{crit} and the final model is established.

The regression analysis of the model was determined with the EXCEL statistical package.

RESULTS

The final model consists of two independent variables, namely: population growth and water losses. Three iterations were conducted until all *p*-values were less than 5%. Economy and water restrictions were eliminated since they were not satisfying the significance level described by α_{crit}. The final regression function established from the regression model (Equation (2)) is given as:

$$y = 7.923 \times 10^7 + 51.465x_1 + 0.851x_2 \quad (3)$$

The constant b_0 describes the water demand if there was no population growth or water loss, which is idealist since such a scenario will never occur. The water demand due to population growth is

Water Practice & Technology Vol 13 No 3
doi: 10.2166/wpt.2018.082

described by the second term and the last term describes the water losses. It is clearly seen from Equation (3) that population causes the largest proportion of water demand.

Figure 2 shows the actual water demand and the water demand developed with the regression function from Equation (3). A close fit was established between the model and the actual water demand data. The jump of the actual water demand curve in 2003 to 2004 is due to restrictions imposed by the city of Cape Town.

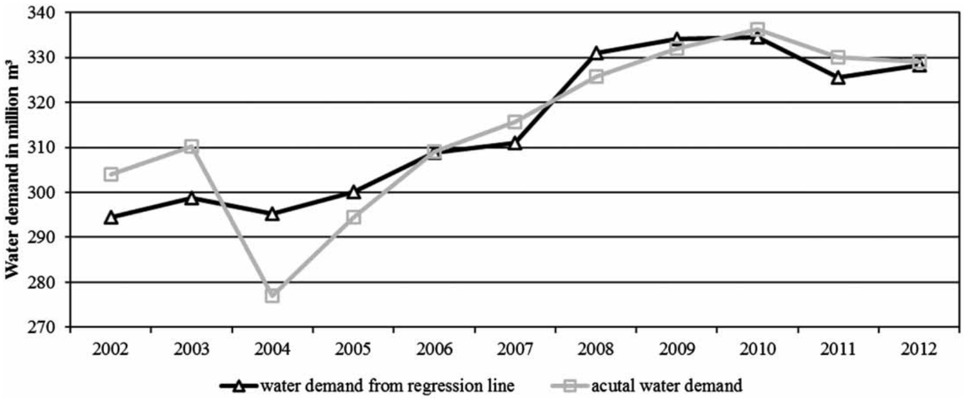

Figure 2 | Regression curve vs. actual water demand 2002–2012 for the city of Cape Town.

Based on the established final regression function (Equation (3)), the future water demand was determined, whereby the average population growth rate was 2.7% based on data of previous years for the city of Cape Town. The future water losses were assumed to stay at the same level of 14% as recorded in the latest data of 2012 (CoCT 2013a). The city of Cape Town expects the losses to stay at the same level because further measures to reduce the water losses cannot be realized due to insufficient funding. A forecast from 2013 to 2029 was considered suitable.

The projected yield of the available sources and projected water demand are shown in Figure 3. The decrease of the system reflects the estimated influence of climate change on the current resources. In this regard, the city of Cape Town assumed that the system yield will be reduced by 5% over the next 25 years due to global warming, starting from 2012 (CoCT 2013c). The projected water demand curve indicates that the water demand increases by $8.12 \times 106\ m^3$ per year, which is about 2.01%.

Based on Figure 3, a new water source will be required by 2021, as indicated by the point of intersection of the two lines. At this point the water demand is equal to the amount of water yielded by the current water sources. Figure 4 shows the projected water deficit by volume. The estimated water deficit of the city of Cape Town from 2022 to 2029 is based on the difference between projected water

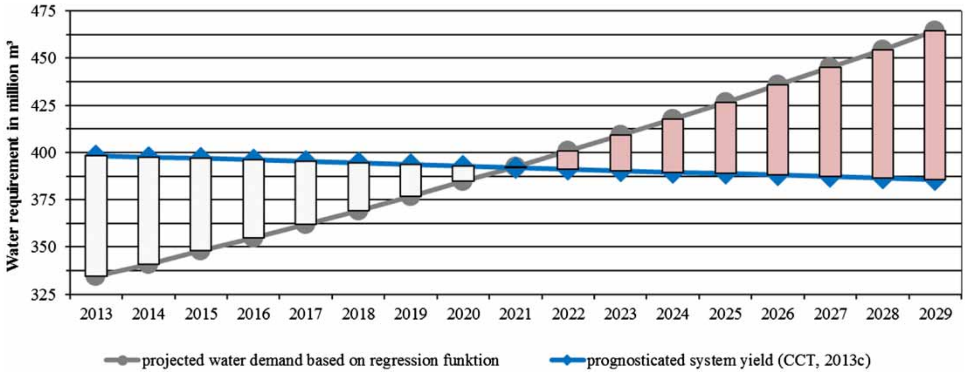

Figure 3 | Projected water supply scenario 2013–2029.

Water Practice & Technology Vol 13 No 3
doi: 10.2166/wpt.2018.082

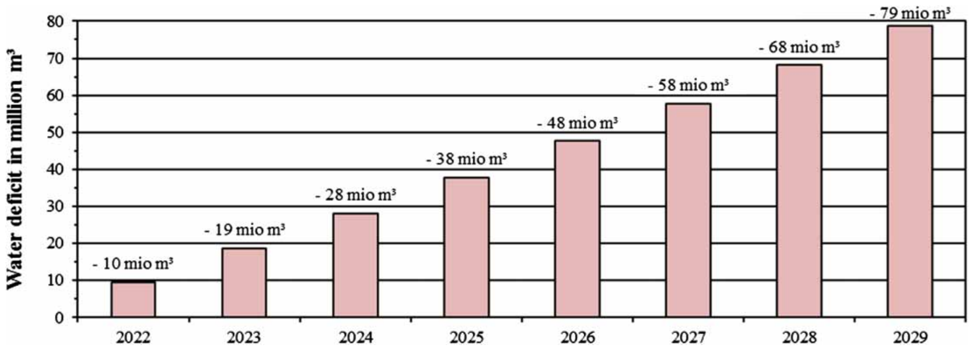

Figure 4 | Projected water deficit 2022–2029.

demand and projected system yield (see Figure 3). However, this trend can change if the other assumptions change like population and water losses.

The city of Cape Town also commissioned studies to determine when a new resource is required to ensure continuity in water supply to the city of Cape Town. Based on the studies of the city of Cape Town, a new water resource would be required between 2016 and 2025, depending on the success of the water conservation and water demand strategy (CoCT 2013c). Although the findings of the city of Cape Town gave a time range when the deficit will occur, overall they tally with the findings of this research.

Based on this study, new water sources would be required by 2021 to ensure continuity in water supply of the city of Cape Town in the future. Since water demand of the city of Cape Town is decisively influenced by population and water losses, water conservation and demand strategies should be adopted accordingly. In addition, most of the water requirement of the city of Cape Town is contributed by surface water, which is also under threat due to climate change and unpredictable rainfall patterns. Therefore, diversification of the water supply sources becomes essential. In this regard desalination of seawater and water reuse are some of the alternative sources that can be explored further. The Table Mountain and Atlantis aquifers have been identified as the groundwater sources with high water supply potential and are still under investigation.

DISCUSSION

Before the regression function was applied, its validity was evaluated and the dependent variables that contributes to the overall model were established. In this regard, a variance analysis (e.g. R, R^2, F-test, residual analysis *p*-value test) was applied.

In order to verify the quality of the model, the value for multiple correlation (R) of the final regression model was analyzed. The correlation coefficient is a measure of the strength of the linear correlation between variables. The value of R was found to be 0.894, which indicates a high positive correlation between the independent and dependent variables (greater than 0.7). This confirms the assumption that a relationship exists between water demand, population, and water losses.

Furthermore, the F-Test, which is another test to clarify the quality and overall significance of the model based on the null hypothesis, was conducted. The calculated F-value is equal to 15.96 and the threshold value of F is equal to 4.46 at 5% significance level. The calculated F-value is greater than the threshold value of F and therefore the model is considered as significant and the null hypothesis is rejected. The significance of F is the *p*-value for the F-Test and is equal to 0.002%. Hence the probability that the null hypothesis is rejected by mistake or erroneously judged to be significant is 0.002%. Since this probability is low, it can be assumed that R^2 is interpreted correctly and the whole model is significant.

Water Practice & Technology Vol 13 No 3
doi: 10.2166/wpt.2018.082

The *p*-value for the variable x_1 is equal to 0.0006 and for the variable x_2 it is equal to 0.0487. In both cases the *p*-value is less that the critical value (<5%). This indicates that both variables are significant in describing the target variable and the null hypothesis is rejected. This in turn means that the population and water losses can be considered as the most suitable subsets of the determined variables to describe the water requirement of the city of Cape Town.

The coefficient of determination R^2 is a measure to describe the quality of the linear approximation of the regression function with regard to the given data points and indicates the correlation between the variables. It measure the quality of the linear approximation of the regression function with regard to given data points. The value of R^2 of the final regression model is equal to 0.80 meaning that there is a good linear approximation of the regression function. This means that 80.0% of the variation in the water demand is explained by the variation in population and water loss, which can be assumed as good linear approximation of the model. Since a small data set has been used for regression analysis, this means that the unexplained deviations of the data points were relatively small.

In order to establish the effects of removing one variable, the adjusted R^2 of the first regression model was compared with the adjusted R^2 of the final regression model. The adjusted R^2 of the first model is equal to 0.851 and the adjusted R^2 of the final model is equal to 0.800. This shows that the adjusted R^2 is reduced only slightly in the process of backward elimination method. Thus the removal of two predictors caused only a minor reduction in fit, indicating that the influence of the removed predictors on the regression model can be considered as negligible. In addition, R^2 of 0.800 means that 80.0% of variation in the water requirement is explained by the variation in population and water losses.

The removal of the economy as one of the predictor variable could also be explained due to the water distribution of the city of Cape Town. More than 50% of the water is for household use, whilst industrial and commercial use accounts for only about 10%. Consequently, private consumption has a significantly higher influence on the water demand development than the economy.

In conclusion, the variance analysis (e.g. R, R2, F-test, residual analysis *p*-value test) revealed that the final model as well the remaining independent variables can be considered as significant. Since the regression model assessed influences on water demand for a period of 10 years only (2002 to 2012), the linear assumptions used for this model cannot be applied to forecast the trends in the longer-term. Hence the need to further improve the model through continuation of data collection.

CONCLUSIONS AND RECOMMENDATIONS

A linear multiple regression model was developed to predict the water demand trend of the city of Cape Town for water resources planning purposes. Regression coefficients and constants were established from historical data. The regression model assessed influences on water demand for a period of 10 years (2002 to 2012).

The final model consists of two independent variables, which are population growth and water loss. Economy and water restrictions were eliminated. To verify the quality of the final regression model the variance analysis (e.g. R, R2, F-test and *p*-value test) was applied, with the result that the final model, as well as the remaining independent variables, can be considered as significant. Based on this study, new water sources would be required by 2021 to ensure continuity in water supply of the city of Cape Town in the future. Since water demand of the city of Cape Town is decisively influenced by population and water losses, water conservation and demand strategies should be adopted accordingly.

Water scarcity is already a global risk and it must be assumed that the pressure on water resources will increase dramatically in the coming decades due to a growing world population, economic

Water Practice & Technology Vol 13 No 3
doi: 10.2166/wpt.2018.082

development and climatic changes as additional stress factor. To ensure a water supply in the future, the sustainable management of water resource becomes vital. Hence, water demand forecasts based on statistical analysis methods will play an important role since significant drivers may be identified. As a result, water conservation and water supply strategies can be adapted accordingly to ultimately enable a sustainable management of water sources.

ACKNOWLEDGEMENT

This paper would not have been published without the support provided by the program committee of the International Young Water Professionals Conferences organized by IWA, WISA and YWP-ZA. www.iwaywpconference.org

REFERENCES

City of Cape Town 2008 *Water Services Development Plan 2008/09 to 2012/2013*, Executive Summary. Draft, South Africa.

City of Cape Town 2011 *Water Services Development Plan 2011/12 to 2015/2016*. South Africa.

City of Cape Town 2013a *Water Conversation and Water Demand Management Strategy. Water Services Development Plan 2012/13 to 2016/15*. South Africa.

City of Cape Town 2013b *Water Resources. Water Services Development Plan 2012/13 to 2016/15*. South Africa.

City of Cape Town 2013c *Update and Additions to Existing Long Term Water Conservation and Water Demand Strategy*. Water Sanitation and Water Demand Management, South Africa.

City of Cape Town 2013d *Economic Statistics*. South Africa.

Farway, J. 2002 *Practical Regression and Anova Using R*. Available at: https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf (accessed 17 September 2013).

Frame, J. & Killick, M. 2005 Integrated water resource planning in the City of Cape Town. *Water SA* **30**(5), 100–104.

OECD 2008 *Cape Town, South Africa, OECD Territorial Reviews*. Available at: https://www.oecd-ilibrary.org/urban-rural-and-regional-development/oecd-territorial-reviews-cape-town-south-africa-2008_9789264049642-en (accessed 9 July 2018).

Otien, F. A. & Ogutu, C. 2005 *Water Demand Management as an Option to Desalination in the Satisfactory Provision of Water in Cape Metropolitan Area*. Department of Civil Engineering, Tshwane University of Technology, South Africa.

Reinboth, C. 2006 *Multivariate Analysis Methods in Marketing Research*. PhD Thesis, Harz University Department of Economics, Germany.

Tadross, M. & Johnston, N. P. 2012 *Climatic Change Projections for Cape Town: Adding Value Through Downscaling*. City of Cape Town, South Africa.