

# **STATISTICAL CHARACTERISATION AND ESTIMATION OF NON-DOMESTIC WATER DEMAND**

HJ van Zyl<sup>1</sup>, AA Ilemobade<sup>1</sup>, JE van Zyl<sup>3</sup> and Y Le Gat<sup>4</sup>

*<sup>1</sup>School of Civil and Environmental Engineering, University of the Witwatersrand,  
Johannesburg*

*<sup>3</sup>Department of Civil Engineering, University of Cape Town*

*<sup>4</sup>Research Unit REBX, Irstea Bordeaux, 50 Avenue de Verdun, 33612 Cestas Cedex,  
France*

Corresponding author: JE van Zyl. kobus.vanzyl@uct.ac.za

## **ABSTRACT**

Estimation of annual average water demand figures is critical for the design and evaluation of water distribution systems. This study evaluated the metered water consumption of more than 67 000 non-domestic consumers in six categories from cities and towns in South Africa. It was found that lognormal distributions provide good descriptions of the annual average daily demand (AADD) distribution in each category. The land use categories Business Commercial, Industrial, Agricultural holdings and Sports & Parks displayed similar median AADDs of between 1.5 and 1.7 kl/property/day. Educational properties used substantially more water (4.7 kl/property/day), while Government & Institutional properties used substantially less water (0.7 kl/property/day). A step-wise regression analyses showed that property size has the greatest impact on water demand for most categories. Finally, a novel statistically based method is proposed for estimating the average AADD of a given number of properties based on an acceptable risk of non-exceedance.

Keywords: water demand; non-domestic; design demand

## **INTRODUCTION**

Accurate estimation of annual average daily demand (AADD) is important for the design of municipal water distribution systems. AADD values normally form the basis for water supply component sizing including pipelines, storage reservoirs, pump stations and water treatment plants. AADD estimates are also important for the sizing of wastewater distribution and treatment systems.

It is useful to distinguish between domestic and non-domestic consumers in the analysis of water demand. Domestic demand is driven by human needs and behaviour,

and can be estimated from historic data and factors such as development level, property size, income and climate.

In comparison, non-domestic demand is highly variable and is primarily driven by the type of activity practiced on the property. For example, a drinks factory and storage depot may have similar property sizes, turnovers and staff numbers, but will consume vastly different quantities of water.

Non-domestic consumers include industrial, commercial and institutional (ICI) users, parks and agricultural holdings (large properties often used for small scale agriculture).

While studies have found relationships between non-domestic demand and parameters such as property size, developed area, economics, water rates and water restrictions (McCuen et al, 1975, Kollar and Brewer, 1977; Stephenson and Turner, 1996; Van Vuuren and Van Beek, 1997; Zhou and Tol, 2005; Reynaud, 2003), these models require information at individual property level. They are less suitable for new areas where the types of non-domestic users that will be established are unknown, or existing areas where the types of developments may change over time.

In recent years, municipal data management tools have become common in South African municipalities, giving researchers access to a large quantity of metered consumption data (Jacobs and Fair, 2012). This data formed the basis for a number of studies on water demand, which identified influencing parameters, investigated typical demand ranges and proposed new design guidelines for the estimation of domestic demands (e.g. Stephenson and Turner, 1996; Van Vuuren and Van Beek, 1997; Jacobs et al., 2004; Husselmann and Van Zyl, 2006; Van Zyl et al, 2008; Griffioen and Van Zyl, 2014), but did not consider non-domestic consumption.

In this study, metered consumption data of more than 67 000 non-domestic properties were linked to various databases and analysed in six categories. The aims of the study were to:

- Estimate the statistical properties and distributions of the data in each category;
- Identify the main factors that influence non-domestic water demand in each the category; and
- Develop a method to estimate the average consumption of a given number of non-domestic properties for a chosen level of confidence.

While detailed guidelines for the water demand of different non-domestic uses are often given in table form, these guidelines are only useful if the types and properties of these users are known. However, this is not the case for new developments and even in existing developments the types of non-domestic consumers are not fixed, but change over time. Thus this study approached non-domestic consumption using a statistical approach, rather than at individual property level.

It is hoped that the result of this study will encourage similar research in other countries, allowing for comparison of non-domestic demand parameters and improved estimation of non-domestic demands.

## **METHODOLOGY**

The data set used in this study is the same as that used by Van Zyl et al (2008) for a study on domestic water demand. It included water meter readings from forty-eight municipal treasury databases totalling more than 2.5 million records in four metropolitan areas (Johannesburg, Tshwane, Ekurhuleni and Cape Town) and 151 other cities and towns. The data is described in detail in Van Zyl et al (2008).

Approximately 438 400 non-domestic consumer records were extracted from the database, and subjected to a number of verification tests to exclude anomalies. The remaining 67 291 records were then separated into six land use categories, namely:

- Business Commercial;
- Industrial;
- Government & Institutional;
- Education, including schools, universities and other educational buildings and their grounds;
- Agricultural holdings, which are large properties serviced by municipal water distribution systems. Agricultural holdings may be used for small scale farming, although alternative sources of water would often be used for this purpose. These properties are also called ‘small farms’ or ‘residential farms’;
- Sports & Parks;

Information on possible factors influencing non-domestic demand was collected from various sources and linked to the demand data to allow correlations to be analysed. This information was obtained from municipal treasury databases (property area, property value and improvements, monthly meter readings), the South African Weather Service (rainfall, average daily minimum and maximum temperatures) and a manuscript on the surface water resources of South Africa (Midgley *et al.*, 1994).

The AADD for each record in the database was estimated from twelve consecutive months of monthly consumption values. This means that seasonal demand variations did not influence the values, and errors due to differences in reading dates or erroneous reading would be small. The meter readings exclude demand from other sources, such as boreholes and grey water.

To ensure the integrity of the data, three data cleaning phases were implemented. In the first phase, records with inconsistent readings or dates (e.g. due to meter replacements or clock-overs) and records with less than 12 consecutive months of data were excluded. In the second phase, records or properties identified as unmetered, vacant, pre-paid, duplicate (including more than one type of land use) or not classified as one of the target non-domestic categories were excluded. In the third phase, stands with unrealistically high or low property sizes or values (municipal valuation of the property) were excluded as detailed in Table 1.

Table 1. Ranges of property areas and values included in this study.

| <b>Land Use Category</b>      | <b>Property area (m<sup>2</sup>)</b> | <b>Property Value (Euro)</b> |
|-------------------------------|--------------------------------------|------------------------------|
| Business Commercial           | 20 - 50 000                          | 1 400 – 7 000 000            |
| Industrial                    | 20 - 50 000                          | 1 400 – 7 000 000            |
| Government &<br>Institutional | 100 - 50 000                         | 1 400 – 1 400 000            |
| Educational                   | 100 - 75 000                         | 1 400 – 1 400 000            |
| Agricultural holdings         | > 5000                               | 1 400 – 700 000              |
| Sport and Parks               | 500 - 100 000                        | 1 400 – 1 400 000            |

The final database contained the following parameters for each record:

- Land use code and suburb;
- AADD, based on 12 consecutive months of consumption;

- Property size;
- Property value (including improvements);
- Urbanisation level (city or town);
- Geographic location (coastal or inland);
- Mean annual precipitation;
- Mean annual evaporation;
- Average minimum daily temperature;
- Average maximum daily temperature.

The database was then used as basis for analysing non-domestic demand characteristics and influencing factors.

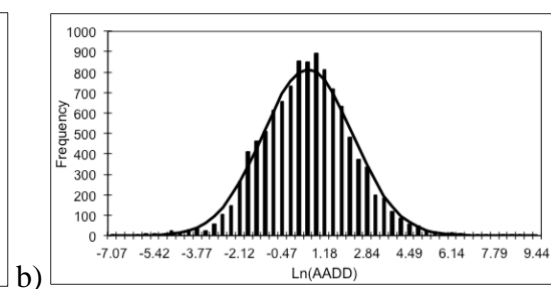
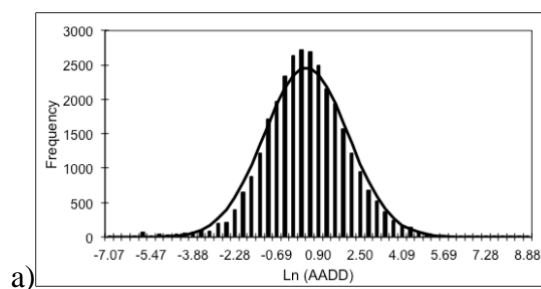
## **NON-DOMESTIC DEMAND CHARACTERISTICS**

In the first phase of the study, the statistical characteristics of the non-domestic demand categories were calculated. It was found that log-normal distributions fit the data remarkably well as shown on the distribution density and cumulative distribution functions in Figures 1 and 2 respectively. Although the lognormal distributions didn't fit the data perfectly (the Kolmogorov Smirnov test rejected normality in the data), they clearly provide reasonable descriptions of the data and are mostly conservative for larger demands.

Table 2 summarises the mean, median and lognormal distribution parameters of the different categories.

Table 2. Statistical descriptors and lognormal distribution parameters of the different non-domestic demand categories

| Category                   | No of data points | Mean AADD<br>(kl/property/day) | Median AADD<br>(kl/property/day) | Lognormal distribution parameters |   |
|----------------------------|-------------------|--------------------------------|----------------------------------|-----------------------------------|---|
|                            |                   |                                |                                  | Mean<br>Ln(kl/property/day)       | Standard deviation<br>Ln(kl/property/day) |
| Business Commercial        | 30 850            | 6.83                           | 1.52                             | 0.4392                            | 1.592                                     |
| Industrial                 | 10 851            | 13.99                          | 1.72                             | 0.5345                            | 1.754                                     |
| Government & Institutional | 12 731            | 5.07                           | 0.67                             | -0.2634                           | 1.399                                     |
| Educational                | 2 146             | 13.29                          | 4.71                             | 1.4251                            | 1.714                                     |
| Agricultural holdings      | 9 612             | 3.56                           | 1.56                             | 0.3631                            | 1.241                                     |
| Sports & Parks             | 1 101             | 9.05                           | 1.67                             | 0.4044                            | 2.033                                     |





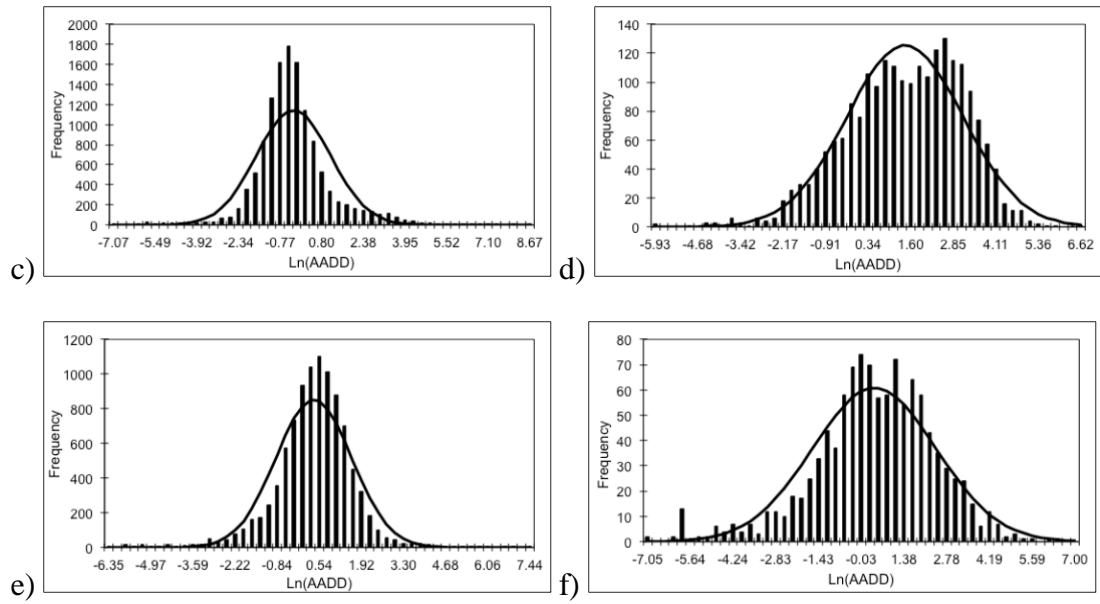


Figure 1. Demand frequency distributions for the non-domestic categories a) Business Commercial, b) Industrial, c) Government & Institutional, d) Educational, e) Agricultural holdings and f) Sports & Parks. In each graph the X-axis represents the natural logarithm of the AADD, and the Y-axis the number of properties.

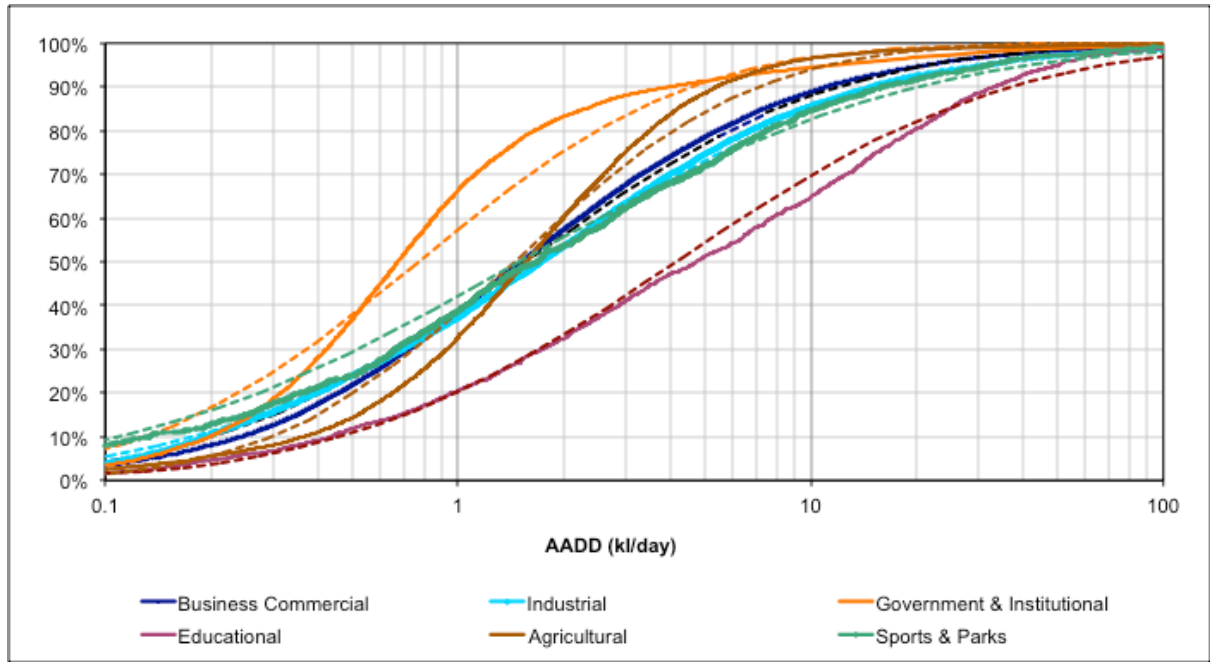


Figure 2. Cumulative distribution functions with lognormal distributions for different non-domestic categories. Solid lines show the data and dotted lines the lognormal distributions.

Due to the long high-end tail of lognormal distributions, the mean demands are considerably larger than the median demands in each category. The median demands for the different categories are compared in Figure 3. The figure shows that Business Commercial, Industrial, Agricultural holdings and Sports & Parks have remarkably similar median demands of between 1.5 and 1.7 kl/property/day.

Figure 2 confirms the similarity of the median values, and further shows that the distributions of Sports & Parks, Industrial and Business Commercial properties are also similar. While agricultural holdings have a similar median demand as the other three categories, it has significantly lower variability.

The two outermost demand distributions are Educational on the higher side, and Government & Institutional users on lower the side, with median demands of 4.71 and 0.67 kl/property/day respectively.

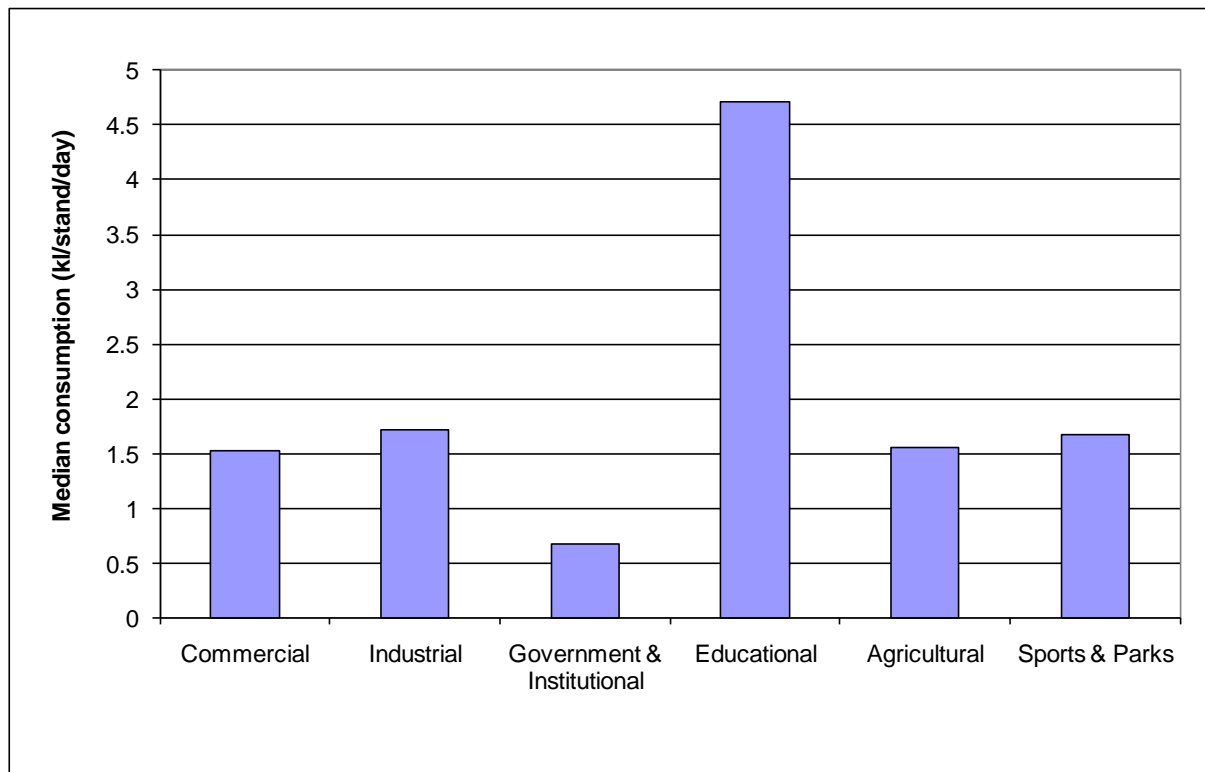


Figure 3. Median demands of non-domestic categories

To remove the impact of property size on the demand distributions, the AADD of each property was divided by its size to obtain the unit demand (measured in  $L/m^2/day$ ). The cumulative distributions are shown in Figure 4 and the mean and median values in Figures 5.

While Government & Institutional users had the lowest consumption per property (Figure 3), Figure 5 shows that they had the highest median demand per unit area. The median demand per unit area of educational users ranks much lower compared to its demand per property, most likely due to the large stands with school playgrounds, parking and sports facilities that are common at these institutions.

An interesting result is that the demand per unit area of agricultural holdings is by far the lowest of all the non-domestic categories.

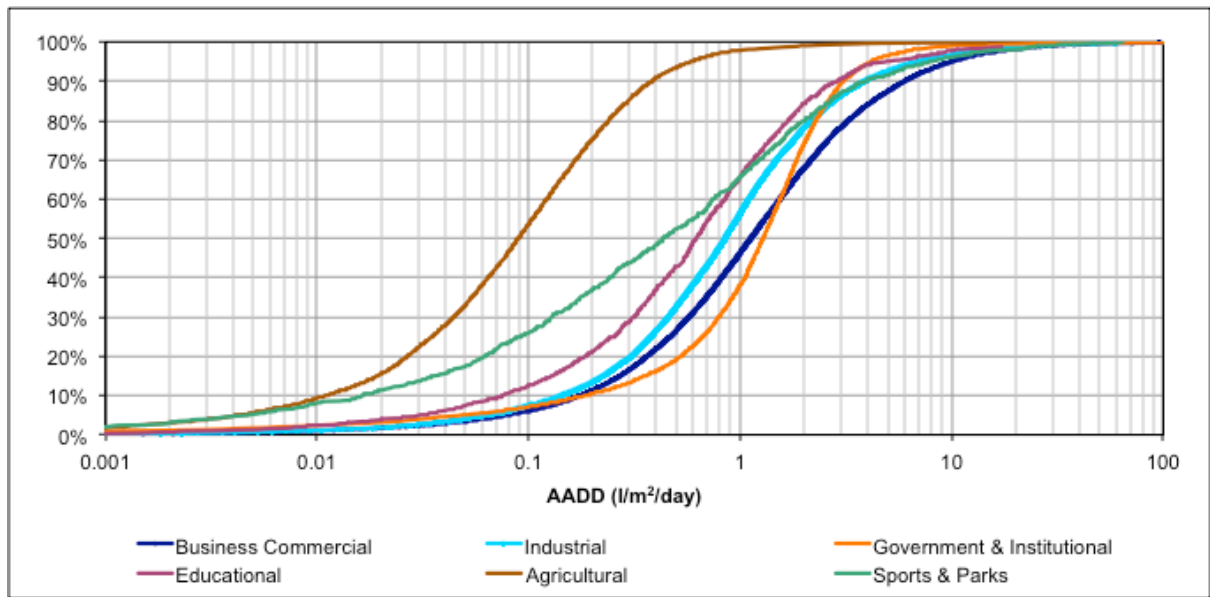


Figure 4. Cumulative distributions of non-domestic demand per unit area.

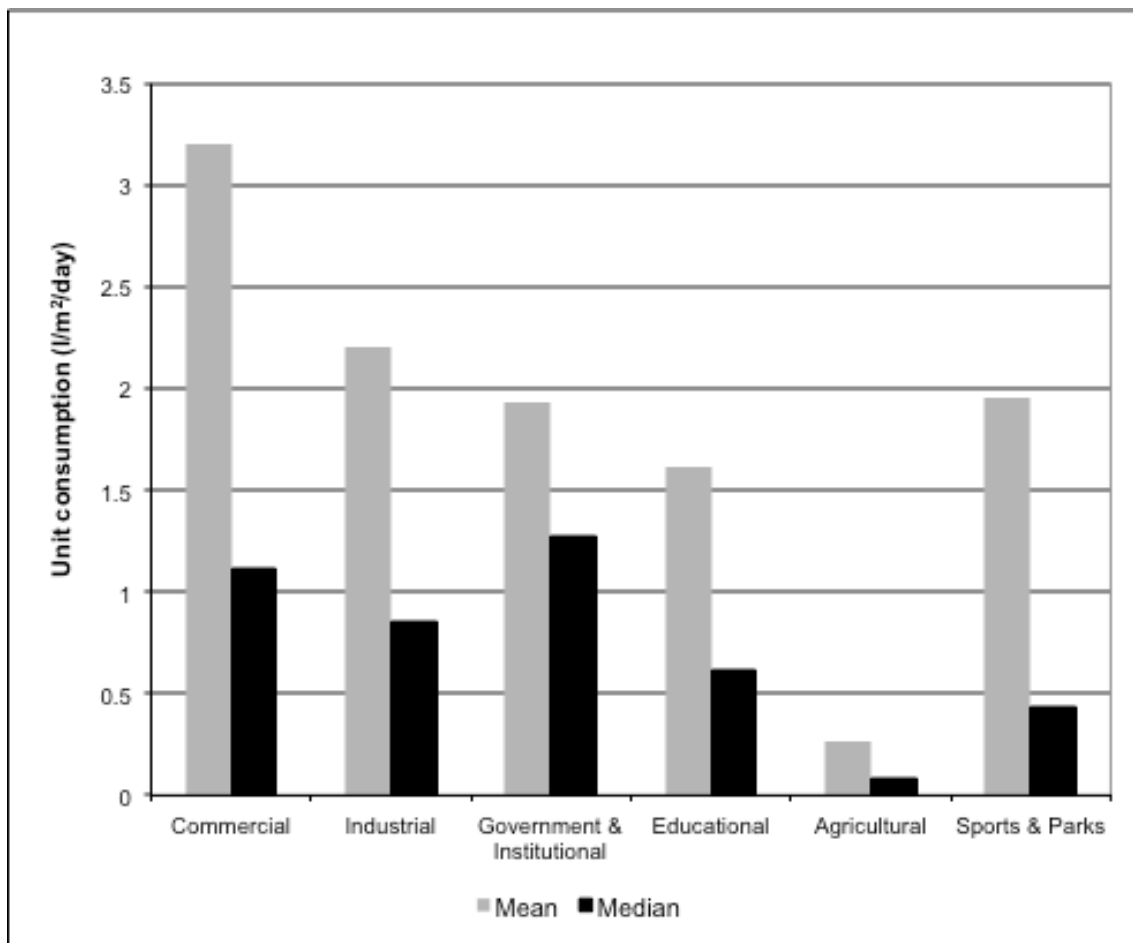


Figure 5. Mean and median demands per unit area for non-domestic categories

## FACTORS INFLUENCING NON-DOMESTIC DEMAND

To analyse the various factors influencing non-domestic water demand, a stepwise multiple regression analysis was performed on the data, using the independent variables listed in the Methodology section, and including the natural logarithms of the property area and property value parameters. Table 3 shows the number of statistically significant parameters identified, as well as the three most important parameters for each category.

Property size emerged as the most important factor in all categories except for Business Commercial and Agricultural holdings. For Business Commercial, property size was the second most important factor (after property value), and for Agricultural holdings, the third most important factor (after property value and temperature).

Table 3. Most important factors influencing non-domestic demand

| Category            | No of data points | No of significant parameters | Most important factors | Cumulative adjusted R <sup>2</sup> |
|---------------------|-------------------|------------------------------|------------------------|------------------------------------|
| Business Commercial | 24 810            | 8                            | Ln(property value)     | 0.230                              |
|                     |                   |                              | Ln(property area)      | 0.283                              |
|                     |                   |                              | Location               | 0.303                              |
| Industrial          | 8 000             | 8                            | Ln(property area)      | 0.313                              |
|                     |                   |                              | Property value         | 0.330                              |
|                     |                   |                              | Location               | 0.345                              |

|                            |       |   |                         |       |
|----------------------------|-------|---|-------------------------|-------|
| Government & Institutional | 8 259 | 8 | Ln(property area)       | 0.336 |
|                            |       |   | Ln(property value)      | 0.373 |
|                            |       |   | Property value          | 0.387 |
| Education                  | 1 447 | 5 | Ln(Property area)       | 0.408 |
|                            |       |   | Development level       | 0.422 |
|                            |       |   | Property value          | 0.437 |
| Agricultural holdings      | 8 895 | 6 | Ln(property value)      | 0.096 |
|                            |       |   | Ave minimum temperature | 0.113 |
|                            |       |   | Property area           | 0.118 |
| Sports & Parks             | 367   | 4 | Ln(Property area)       | 0.137 |
|                            |       |   | Ave minimum temperature | 0.176 |
|                            |       |   | Development level       | 0.191 |

### MODELLING THE MEAN DEMAND OF $k$ STANDS

The final aim of the study was to develop a model to predict the mean demand of  $k$  stands of a category at a given level of confidence. The theoretical question of finding a high quantile for a given probability level (95% was used in this study) of a sum of a given number of random variables seems to have received little attention in the

statistical literature in the case where the distribution of the random variables is unspecified. This fact was pointed out early by Watson and Gordon (1986), who investigated the relationship between the quantiles of a sum of two independent random variables and those of its components.

The problem was recently addressed by Klass and Nowicki (2010), but the method seems to only provide loose bounds not practically usable for the purposes of this study. Extensive work is reported by Nadarajh (2008) in the case where the distribution of the sum is parametrically defined, but the results related to the log-normal distribution cannot be applied to the consumption data of this study, for which the lognormal fit appears to be skewed by the weight of the upper tail of the empirical distributions.

Thus it was decided to adopt a heuristic approach based on discarding the highest values of the tail of the distribution, and then fitting an approximate model to the remaining cumulative distribution. In practical terms this means that a design based on these values will not cater for properties with exceptionally high demands.

The model that was used has the form:

$$q_k = \mu + \frac{q_1 - \mu}{\sqrt{k}} \quad \dots(1)$$

Where  $q_k$  is the design demand per property,  $\mu$  the expected value of the distribution,  $q_1$  the chosen percentile of the data, and  $k$  the number of properties.

The choice of the analytical form of the equation was guided by the following considerations:

- Compliance with the limit condition  $q_k = q_1$  for  $k = 1$ , i.e. when only one property is selected, the design value should be at the chosen percentile of the original data.

- A decrease of the design value proportional to  $1/\sqrt{k}$ , by analogy to the decrease in the standard deviation.

The threshold value was determined by trading off two objectives: the first to discard as few data points as possible, and the second to find a suitable fit between the model and data. To achieve this, a data set was first truncated at a given demand threshold, and the 95 percentile of the demand then determined as a function of the number of properties through Monte Carlo sampling. The model parameters giving the best fit to the 95-percentile curve were then estimated.

This model was only applied to the Business Commercial, Industrial and Agricultural holdings categories since, unlike the other categories, these stands are normally grouped together. The demand threshold and model parameters determined for the three demand categories are summarised in Table 4, and the functions and data are shown in Figure 6 (for a certainty of 95 % that they will not be exceeded).

Table 4. Model parameters for the 95 percentile curves

| <b>Category</b>          | <b>Demand<br/>threshold<br/>(kl/property/day)</b> | <b>Fraction<br/>of data<br/>excluded</b> | <b><math>q_1</math><br/>(kl/property/day)</b> | <b><math>\mu</math><br/>(kl/property/day)</b> |
|--------------------------|---|--|---|---|
| Business<br>Commercial   | 50  | 1.7 %                                    | 16.340  | 3.806   |
| Industrial               | 80  | 1.8 %                                    | 23.032  | 5.179   |
| Agricultural<br>holdings | 20  | 1.3 %                                    | 6.950   | 2.276   |



## PROPOSED DESIGN GUIDELINE

The models developed above provide a useful design guideline for Business Commercial, Industrial and Agricultural holdings developments where stands are grouped and the details of consumption are not known in advance.

The design mean demand for developments, based on a 95 % certainty of non-exceedance, may be estimated using the equations:

$$\text{Business commercial:} \quad q_k = 3.806 + \frac{12.534}{\sqrt{k}} \quad \dots(2)$$

$$\text{Industrial:} \quad q_k = 5.179 + \frac{17.853}{\sqrt{k}} \quad \dots(3)$$

$$\text{Agricultural holdings:} \quad q_k = 2.276 + \frac{4.674}{\sqrt{k}} \quad \dots(4)$$

Where  $q_k$  is the design AADD in kl/property/day and  $k$  the number of properties.

The design models, based on Monte Carlo sampling, are given in Figure 6. The models provide a good description of the demand behaviour, except for Commercial areas with less than 6 stands, where it is advisable to use the data points themselves.

As an example, consider a new industrial area with 50 stands. From Figure 6, the design demand can be estimated as 7.7 kl/property/day, or a total AADD of 385 kl/day for the 50 stands. This is significantly higher than the category mean demand of 5.2 kl/property/day, to accommodate the likelihood of a higher concentration of high-demand stands in the area.

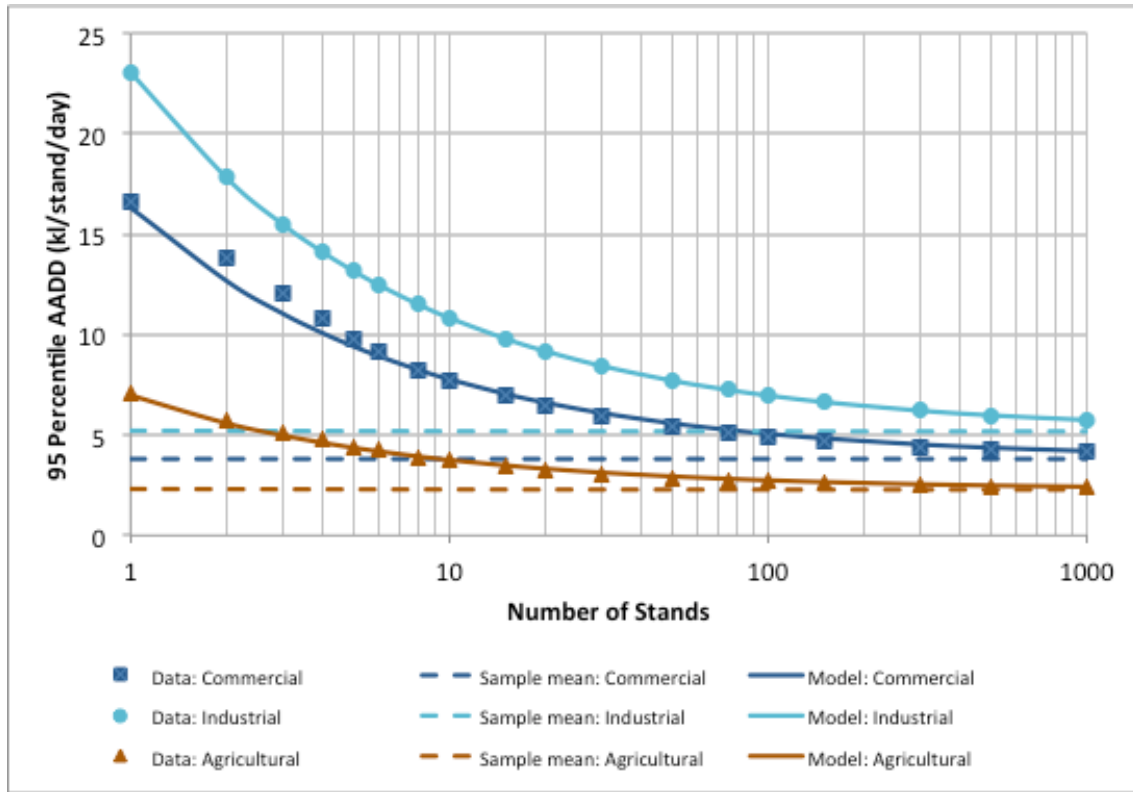


Figure 6. Proposed design demand guidelines for Business Commercial , Industrial and Agricultural Holdings properties

## CONCLUSION

This study analysed six categories of non-domestic demand in South Africa based on metered consumption data for a large number of stands. Basic statistical descriptors and distributions are reported and compared. The lognormal distribution was found to provide a good description of the demand distribution in all categories. A step-wise regression analyses was used to analyse various influencing factors, and property size was found to be the most important factor.

From the results obtained, guidelines for estimating water demand based on a 95 % certainty of non-exceedance were developed for Business Commercial, Industrial and Agricultural property areas. It is hoped that this study will encourage comparative

studies in other countries leading to a better understanding of the range of demand parameters and improved design guidelines for non-domestic water demand.

## **ACKNOWLEDGEMENTS**

The authors acknowledge financial support for the project provided by the Water Research Commission (Project K5/1525) and water demand data supplied by GLS Consulting.

## **REFERENCES**

- CSIR, 2003. *Guidelines for Human Settlement Planning and Design*, 3<sup>rd</sup>, Council for Scientific and Industrial Research, Pretoria, South Africa.
- GLS, 2007. *SWIFT*. Available from:  
<http://www.gls.co.za/Gls/pages/swift/overview.html>.
- Griffioen, M.L., and Van Zyl, J.E. 2014. Modelling water demand on a suburb level, *Journal of the South African Institution of Civil Engineering*, 56 (1), 63-68.
- Husselmann, M.L., and Van Zyl, J.E. 2006. Effect of property size and income on residential water demand. *Journal of the South African Institution of Civil Engineering*, 48 (3), 12-16.
- Jacobs, H. E., Geusteyn, L.C., Loubser, B.F. and Van Der Merwe, B. 2004. Estimating residential water demand in Southern Africa. *Journal of the South African Institution of Civil Engineering*,. 46 (4), 2-13.
- Jacobs, H.E. and Fair, K.A. 2012. Evaluating a tool to increase information processing capacity for consumer water meter data, *South African Journal of Information Management*, 14 (1), Art. #500, 7 pages.
- Klass, M.J. and Nowicki, K. 2010. Uniformly accurate quantile bounds for sums of arbitrary independent random variables. *Journal of Theoretical Probability*, 23, 1068–1091.
- Kollar, K.L. and Brewer, R, 1977. The Impact of Industrial Water Use on Public Water Supplies. *Journal American Water Works Association*, 69 (9) 468-473.

- McCuen, R. H., Sutherland, R. C. and Kim, J. R., 1975. Forecasting Urban Water Use: Commercial Establishments, *Journal American Water Works Association*, 67 (5), 239-244.
- Midgley, D.C., Pitman, W.V. and Middleton, B.J., 1994. *Surface Water Resources of South Africa*. WRC Report No. 298/1/94. Water Research Commission, Pretoria, South Africa.
- Nadarajh, S., 2008. A review of results on sums of random variables. *Acta Applicandae Mathematicae*, 103, 131–140.
- Reynaud, B., 2003. An econometric estimation of industrial water demand in France. *Journal of Environmental and Resource Economics*. 25 (2) 213-232.
- Stephenson, D. and Turner, K., 1996. Water demand patterns in Gauteng. *Institute of Municipal Engineering of Southern Africa*. 21 (1) 11-16.
- Van Vuuren, S.J. and Van Beek, J.C., 1997. *Herevaluering van die Bestaande Riglyne vir Stedelike en Industriële Watervoorsiening gebaseer op Gemete Waterverbruik Fase 1: Pretoria Voorsieningsgebied*, WRC Report No. 705/1/97, Water Research Commission, Pretoria, South Africa.
- Van Zyl, H.J., Ilemobade, A.A. and Van Zyl, J.E., 2008. An improved area-based guideline for domestic water demand estimation in South Africa. *Water SA* 34 (3), 381-392.
- Watson, R. and, Gordon, I., 1986. On quantiles of sums. *Australian Journal of Statistics*, 28 (2), 192–199.
- Zhou, Y. and Tol, R. S. J., 2005. *Water Use in China's Domestic, Industrial and Agricultural Sectors: An Empirical Analysis*. Working paper-67, FNU University, Hamburg, Germany.