Preprint fo paper accepted for publication in Sustainable Cities and Society

Fees and governance: Towards sustainability in water resources management at schools in post-apartheid South Africa

M. J. Booysen [a], Buddhi Wijesiri [b, c, *], Cheroline Ripunda [a], Ashantha Goonetilleke [c]

[a] Department of Electrical and Electronic Engineering, Stellenbosch University, Stellenbosch 7600, South Africa

[b] College of Chemistry and Environmental Engineering, Shenzhen University, Shenzhen 518060, China

[c] Science and Engineering Faculty, Queensland University of Technology (QUT), GPO Box 2434, Brisbane, Qld 4001, Australia

Highlights

- Affluence and self-governance drive water usage in the schools assessed
- Non-fee-paying schools use more than fee-paying schools
- For non-fee-paying schools, 50% of schools responsible for 85% of use
- Number of students and number of educators do not affect usage rate
- Usage in hours before and after school and Saturdays proxy heavy use

Abstract: Water scarcity is increasingly staking a claim next to energy as a threat to the sustainability of large cities, especially in developing countries with limited resources. The recent crisis brought on by Cape Town's "Day Zero" drought created the impetus to expand on existing research on water demand management to include analysis of school usage patterns and key determinants thereof. With the effects of apartheid still visible in society and in school infrastructure coupled with the high water usage rates at schools, this paper evaluates the impact of school affluence (whether it is fee-paying or not, and self-governing or not) on water usage. We find that poor schools use substantially more water, partially because of poor maintenance, with mean water efficiencies of poor schools around 50% and 80% for affluent schools. Bayesian models were used to further determine which characteristics of a school are good proxies for the higher usage to help administrators and policy makers in the resource constrained educational environment. In addition to the obvious impact of maintenance, the results point an incriminatory finger at early morning-school usage, early afternoon usage, and Saturday usage.

Keywords: Community affluence; Schools water usage; Sustainable water management; Water demand modelling; Water equity, Water Scarcity

1.0 Introduction

Water shortages are increasingly reported compromising the sustainability of several large cities and regions worldwide. The unpredictability and extremity of climate change have further intensified the gravity of limited water supplies (McDonald et al., 2014; Srinivasan et al., 2017; Wagener et al., 2010). This problem is particularly salient in developing countries that are characterised by rapid population growth, high rates of urbanisation, and management challenges (Muller, 2018; Ziervogel, 2019). For example, Cape Town recently experienced its worst drought in over 100 years and was declared a disaster area with the so-called "Day Zero" an imminent threat (Enqvist and Ziervogel, 2019).

It is known that service providers and users require accurate and timely usage of information and billing to influence prudent user behaviour and to effectively predict and manage demand. Despite this need, municipalities in developing countries struggle to capture and report on water usage, often relying on estimates of water usage for the billing process (Booysen et al., 2019a; Booysen et al., 2019b; Parks et al., 2019). The result is that users often receive actual billing information two months or more after usage, resulting in undetected leaks and broken feedback information loops. Moreover, in some cases this is further exacerbated by four separate entities, respectively, with responsibility for using the water, maintaining the infrastructure, sourcing the money, and paying the bill. This paper expects to contribute by addressing the issue of reliable water usage data.

There has been a substantial amount of research dedicated to urban water demand management, which is particularly essential in developing countries as they often suffer from high rates of urbanisation. The majority of existing research on urban water demand management have focused on the residential sector, for example, demand forecasting (Adamowski et al., 2012; Bougadis et al., 2005; Donkor et al., 2012; Ghiassi et al., 2017; Ren and Li, 2016), demand modelling (Gurung et al., 2014; Jacobs and Haarhoff, 2004), general demand management (Kenney et al., 2008), and water usage management interventions (Datta et al., 2015; Dernoncourt and Lee, 2016; Fielding et al., 2012). There is limited research on the water demand in the non-residential or educational sectors despite the fact that these sectors can be high water consumers (Sánchez-Torija et al., 2017). Moreover, although historic water usage data has been used in several studies to model water usage patterns, there are several influential factors including socio-economic, political and climatic variables that have not been specifically taken into consideration (Botai et al., 2017; Donkor et al., 2012; Enqvist and Ziervogel, 2019; Muller, 2018; Scheba and Millington, 2018).

In light of Cape Town's "Day Zero" threat, the Western Cape Education Department (WCED) stated that schools are of primary importance as the Province struggled with the drought and to keep schools from closing due to the water shortage (WCG, 2017). Since the schools are responsible for their own water bills, albeit indirectly in some cases, any money that is unnecessarily spent on water reduces the already constrained resources available for education-related expenses. A study by Ripunda and Booysen (2018) highlighted the severity of water wastages in the Province's schools, by showing that a single primary school used as much as 35 kL per day, the equivalent of more than 100 households (Booysen et al., 2019b). The study further demonstrated that significant savings are possible through raising awareness and influencing water usage behaviour. A follow-on maintenance campaign by Booysen et al. (2019a) further demonstrated that even greater savings could be achieved through "quick-and-dirty" inexpensive maintenance at these schools. However, with more than 1,600 schools in the province (more than 23,000 in the country), and with the limited budgetary and managerial resources available, knowing where to focus attention remains a challenge without reliable higher frequency metering data.

Accordingly, we explored the non-residential sector of urban water demand in a developing city context. Specifically, we identified the general trends in water usage by schools in Cape Town, South Africa. Given the scars left by apartheid and severe inequality, we evaluated the influence of a school's affluence, revenue stream, and governance locus of control on their water usage, and in order to identify key drivers in relation to water usage. The results are expected to empower policy makers to focus their attention on the critical areas that drive high usage. Moreover, the results can be used to improve sustainable water management by reducing water usage and the related expenses.

2.0 Materials and Methods

2.1 Case study description

2.1.1 Education system in South Africa

The South African education system prior to the country's first democratic election in 1994 was both, unjust and biased, and the political system was one of totalitarianism with regards to school management. Because of this, after the end of the apartheid regime, the Education Department established several policies aimed at transforming the education system to be just and fair to all South Africans (Dalgleish et al., 2007; Engelbrecht and Harding, 2008; Government_Gazette, 1996; Longueira, 2016). Considering this, the Education Department created two main policies. The first was the SASA (South African Schools Act), which was created to establish committees that would be responsible for the general management of schools. The second was the NNSSF (National Norms and Standards for School Funding), which stipulated the governmental funding for each school according to its socio-economic status.

The SASA of 1996 (Government_Gazette, 1996) aimed to involve communities and relevant stakeholders in the day-to-day management of schools. This was achieved by establishing committees that are responsible for the overall governance of schools. These committees are referred to as School Governing Boards (SGBs) and are made up of educators, parents and learners in the case of secondary/high schools. Thus, the introduction of SGBs brought about shared responsibilities in terms of school governance in South Africa, by involving communities in their own upliftment through improved education. In the name of a fair and just system, the SASA defined the responsibilities of SGBs based on the socio-economic status (SES) of each school. Consequently, two types of schools were defined; termed Section 20 (S20) and Section 21 (S21) schools. For S20 schools, those with lower SES, the government is responsible for buying school material, paying utility bills, and performing maintenance. Section 21 schools on the other hand are allocated funding, from which the SGBs purchase all school materials, pays utility bills and perform their own maintenance. Therefore, SGBs of S21 schools have added responsibility and directly control school fund expenditure. Moreover, SGBs are mandated to augment state funding by implementing either school fees, in the case of some schools, or undertaking fund-raising programmes. These fund-raising programmes include renting out the school grounds to churches and other community groups for a fee. These fund-raising programmes indicate that the allocated governmental funding is not sufficient to sustain general school operations.

The National Norms and Standards for School Funding (NNSSF), which was established in 1998, stipulates how much governmental funding each school receives (Swartz, 2009). Governmental funding is allocated to schools based on their quintile ranking, which divides schools into five groups according to their socio-economic status (Engelbrecht and Harding, 2008; Motala, 2015). Schools in quintiles 1 to 3 are classified as less affluent schools based on their SES. These schools receive higher governmental funding than schools in quintiles 4 and 5 and do not charge fees. For quintile 4 and 5 schools, governmental funding is significantly less and schools can charge school fees to augment their funding. The aim of the system is to remedy the inequality and inequity caused by the apartheid system, by increasing governmental funding to schools with a lower SESs in order to provide better opportunities to previously disadvantaged learners through a better education (Longueira, 2016). Therefore, this policy is expected to create better opportunities for learners that were previously disadvantaged by the old regime.

2.1.2 Water supply and use in South African schools

Water supply within South African schools is unreliable, especially for schools in poorer communities. Currently, South Africa has a total of 23,589 schools. From these, 452 schools were recorded as not having water supply, while another 4,773 have unreliable supply and more than 4,500 still use pit latrines (DBE, 2015).

Western Cape Education Department (WCED) water usage database indicates that four methods are used for recording a school's monthly water usage reading: physical readings by the school; readings by the municipality; automatic estimation; and re-estimation if over estimation occurred. From these, the two commonly used methods are automatic estimation and collection by municipality. The issue was particularly evident in the database of the WCED on schools' water usage data, a snapshot of which can be found in the Supplementary Information. The majority of schools in the Western Cape had several months with no recorded water meter readings in the database, of which the worst case was a school that had no recorded data for 10 months in 2017.

Furthermore, several schools reported that water bills are only issued every two months despite the fact that several of these are responsible for directly settling their own water bills. Consequently, schools are unable to effectively monitor or track their water use patterns. This delayed feedback also makes it difficult for schools to detect and deal with maintenance issues, such as leaks, in a timely manner.

2.2 Data collection

Accurate water usage data is essential for building water demand models that can generate reliable water usage estimates to be used for planning by utilities companies (Bakker et al., 2013; Ferraro and Price, 2013; Ghiassi et al., 2008). Although past studies have utilised several data sources, for example, municipal data, they are known for being inaccurate and unreliable (Datta et al., 2015; Ferraro and Price, 2013). Another data source frequently employed is smart water meters (Fielding et al., 2012; Gurung et al., 2014; Liu et al., 2016). However, smart water meters have only recently been introduced in South Africa, and not yet for schools. Therefore, there is limited or no access to high frequency and accurate long-term water usage data.

This study employed data sets from two different sources. One was a data set of 242 schools located in the Western Cape, obtained from the database emanating from a water-saving campaign of approximately 350 schools during Cape Town's drought, run as a private-public partnership with universities, government, and almost 100 corporate entities (www.schoolswater.co.za/) (Booysen et al., 2019a). Using a smart water meter called a Dropula, water flow was reported in real time to an online platform.

The second dataset was from the WCED, which had details of all the schools within the province. Among the variables were the number of students and educators in each school, whether the school is S20 or S21, and fee-paying or not. These are the variables used in this study.

2.3 Data analysis

2.3.1 Data pre-processing

The data set from the Dropula device was made up of minutely water usage data for each participating school. This data set of 242 schools was first screened based on the continuity of water usage. As such, water usage over a continuous period of at least 720 hours (30 days) or more was considered. This reduced the data set to 163 schools, of which the schools that had zero water usage were eliminated. Accordingly, the final data set included 156 schools.

From this data set of 156 schools, several variables were identified and used for the data analysis. The temporal identifiers were chosen based on observed and anecdotal evidence of school water usage patterns. Some examples are: (1) it was observed that some schools double as church buildings on Sundays, which will affect their Sunday usage; (2) some poorer schools seemed to have maintenance problems, which was linked to nightly flow; (3) some schools have feeding schemes, which will affect the lunch-time water usage; (4) some affluent schools have sporting activities on Saturdays, which will increase Saturday usage; (5) some schools have people living on the property or community members who do not have water supply, may use water from the school's supply during evening and early morning periods; (6) some affluent schools have after school hour music and drama lessons; and (7) some poorer schools have adult education programs in the evenings.

Table 1 captures the data types and classification. Daily usage for each school was separated into weekday, Saturday and Sunday usage. The weekday usage was further divided into different times of the day, which were chosen based on school operating times and activities. These were before school hours, during school hours, extra mural activity hours in the afternoon, after school early evening hours and midnight hours.

Table 1. Data types and classification.

Variable	Data type	Classification	
		Primary	Secondary
Water usage	Quantity (L/hour)	V_w: Weekdays	V_{0508}: 05:00 – 08:00
			V_{0814}: 08:00 – 14:00
			V_{1417}: 14:00 – 17:00
			V_{1722}: 17:00 – 22:00
			V_{2205}: 22:00 – 05:00
		V_{sa}: Saturdays	
		V_{su}: Sundays	
		V_t: Total	
St: Number of students	Quantity		
Edu: Number of educators	Quantity		
Fees: Fees charged	Yes/No		
S21: Self-governance	Yes/No		

2.3.2 Selection of analytical technique

The main imperative in selecting the appropriate analytical technique was the integration of both, quantitative and qualitative variables (see Table 1 above), which were identified to influence schools' water usage. In this context, Bayesian Networks (BNs) modelling has proven to be effective in relation to a range of environmental systems/processes modelling (Bonotto et al., 2018; Borsuk et al., 2004; Liu et al., 2018; Maeda et al., 2017; Martín de Santa Olalla et al., 2007; Rigosi et al., 2015; Ticehurst et al., 2007; Wijesiri et al., 2018). In fact, Bayesian statistical methods have gained relatively little attention, although they have been used for scenario-based water demand modelling. These methods combine the theory of probability and deductive reasoning to manage uncertainty in data.

The BNs modelling facilitates developing interdependencies between variables using the current knowledge of the problem, and their *Markov Property* (i.e. each variable depends only on its immediate parent variables) and overcomes the *curse of dimensionality* when dealing with small data sets (Scutari, 2009). A detailed discussion on BNs modelling is provided in the Supplementary Information. Accordingly, BNs modelling was employed in the current study to understand the interdependencies between influential factors of water demand in the schools. The modelling outcomes were then used to assess the significance of the state of affluence of schools compared to other factors.

3.0 Results and Discussion

3.1 General trends in water usage by schools

Table 2 summarises the data captured for different scenarios, and summarises the number of schools in each scenario. From the 156 schools investigated, 27 are in Scenario 1, 44 in Scenario 2, 12 in Scenario 3, and 73 in Scenario 4. In summary, this translates to 73 affluent schools and 83 less affluent schools in the dataset.

Table 2. Summary of schools in the dataset.

Scenario			**Number of Schools (%)**	**Description**
No.	Fees	S21		
1	No	No	27 (17)	Parents don’t pay, school not self-governing
2	No	Yes	44 (28)	Parents don’t pay, school self-governing
3	Yes	No	12 (8)	Parents pay, school not self-governing
4	Yes	Yes	73 (47)	Parents pay, school self-governing

The results in Fig. 1(a) show a drastic difference in flow rate for each school over all hours (V_t) from Scenarios 1 – 4. The medians and means are incrementally less for each scenario, with the Scenario 4 mean, 189 L/hr, only 40% of the Scenario 1 mean at 468 L/hr. We then investigated the source of the difference between the groups by individually evaluating the periods in Fig. 1(b) to Fig. 1(f).

As expected, the highest flow rate for each scenario occurs during school hours (8:00 to 14:00), with disparate means of 709 L/hr, 536 L/hr, 331 L/hr and 364 L/hr, respectively, for the four scenarios. There is large variance in the flow rates of the two non-fee-paying schools, with the fee-paying S21 scenario (Scenario 3) using similar amounts of water as the fee-paying S21 schools (Scenario 4), but with substantially less variance, which may be because of the small number of schools in Scenario 2.

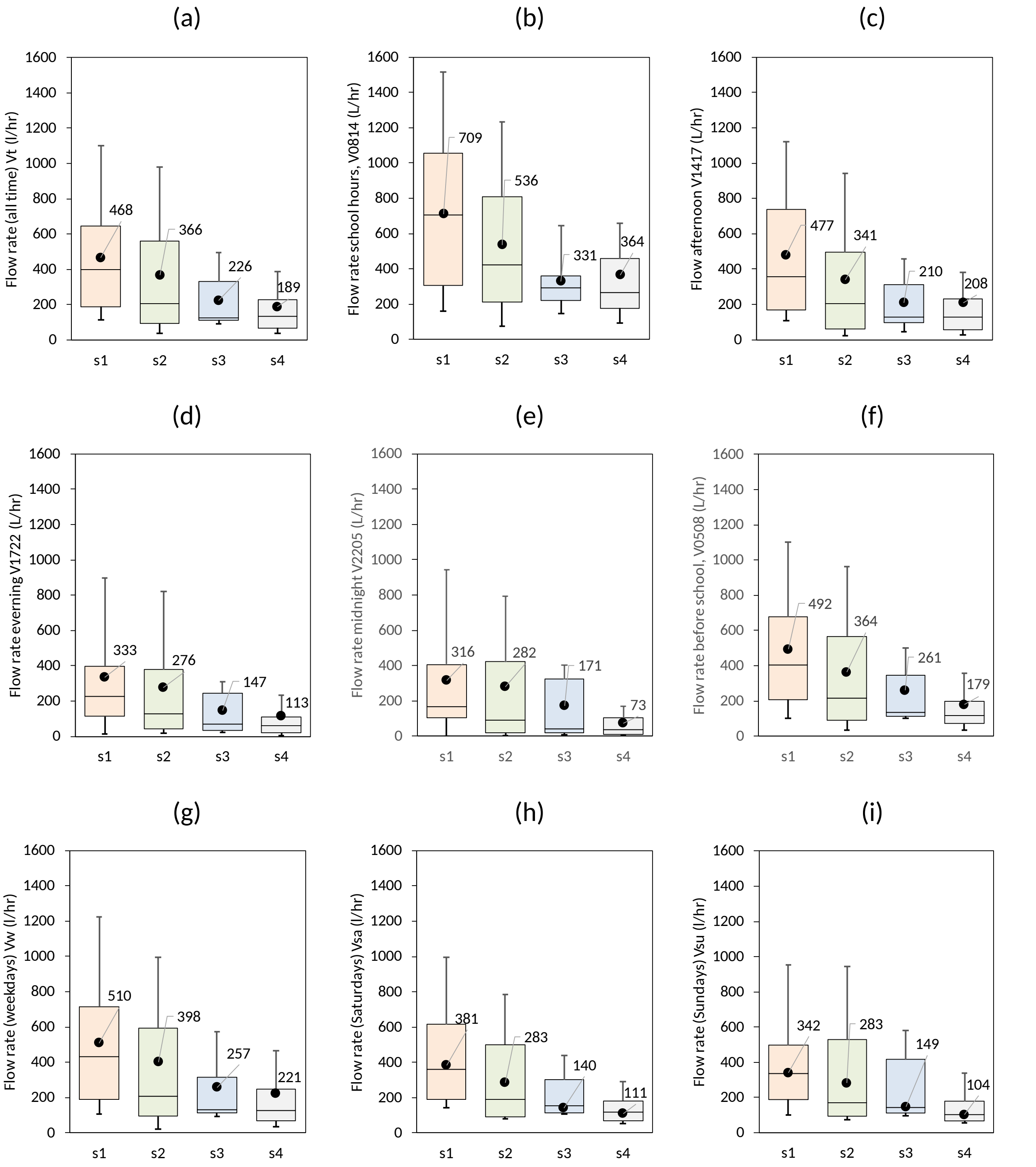

Fig. 1. Temporal distributions of water usage during specified times for Scenarios 1 –4: (a) Total – V_t; (b) School hours – V_{0814}; (c) Afternoon – V_{1417}; (d) Evening – V_{1722}; (e) Midnight hours – V_{2205}; (f) Before school – V_{0508}; (g) Weekdays – V_w; (h) Saturdays – V_{sa}; (i) Sundays – V_{su}.

In the early afternoon, shown in Fig. 1(c), the usage for all scenarios is less than during school hours, as expected. The schools in Scenario 1 still use substantially more than any of the other groups, and Scenario 2 more than Scenario 3 and Scenario 4. Interestingly, not a single school in Scenario 1 drops below 110 L/min during this time. Except for a few apparently wasteful schools in Scenario 3, the median and mean for Scenario 3 and Scenario 4 are virtually on par. Since the extramural activities occur during these hours, it is clear that the flow rates during extramural activities is substantially less than that during school hours, which is also the case for Scenario 4.

The early evening timeslot, Fig. 1(d), mostly exhibits the same patterns and the midnight hours, with an interesting exception – the mean flows during these hours are lower than the mean midnight hours for the two middle scenarios. The medians, however, are lowest for the midnight hours for all four scenarios. All three less affluent schools exhibit a large difference between the median and 75th percentiles, indicating that the "bad half" of those schools have caused this apparent increase. In fact, for the non-fee-paying schools, the top 50% of schools are responsible for 85.2% of the total use. This may be indicative of either usage after 22:00 or usage before 05:00 for those schools, possibly by the surrounding poorer communities.

As expected, the lowest mean flow rate for each scenario occurs between 17:00 and 05:00hrs, with the lowest median flows occurring during midnight hours (22:00 to 05:00). The flow during the hours before school, 05:00 to 08:00, mostly reflects the natural transient from the midnight hours to the school hours, as staff and students arrive during that time.

The midnight flow – means of 316 L/hr, 282 L/hr, 171 L/hr and 73 L/hr – is expected to be zero or close to it, and unexpected flow indicates anomalies, for example, taps left open, stuck toilets, or indicates leaks. Considering that these anomalous flow rates are likely to also be present during the day, we can perform a rough estimate of the intentional water usage as the difference between the school-hour use and the midnight-hour use, resulting in 393 L/hr, 254 L/hr, 160 L/hr, and 291 L/hr for the four scenarios, which gives approximated school-hour efficiencies, calculated by $\eta_{V_{0814}} = (V_{0814} - V_{2205})/V_{0814}$, of 55%, 53%, 48% and 80%, respectively, demonstrating a stark difference between the affluent and poorer schools. However, importantly, when taking the medians, rather than means, the efficiencies are much closer, resulting in 76%, 85%, 79%, and 86% respectively, which indicates that the substantial inefficiencies are largely caused by a the few errant schools, and also shows a trend of higher efficiency for self-governed (S21) schools (Scenarios 2 and 4). These results are also visible when considering weekday (vs. weekend) volume used as a proportion of total use, which results in 78%, 82%, 78%, 84 % for the means and 81%, 86%, 83%, and 88% for the medians, further demonstrating better efficiency for the self-governing S21 schools. The usage on weekdays is double that of weekend days for the affluent s4 schools, but only approximately a third more for the poorer Scenario 1 schools, further underlying that poor maintenance may be at play. The weekend distributions also belie that only the poorest schools, Scenario 1, have substantially more flow on Saturdays than on Sundays, further potentially pointing to community usage during those days.

Another perspective on inefficiencies is given by the volume used in the school hours as a ratio of the total volume, calculated by $\eta_{V_t} = V_{0814}/V_t$, which results in 38%, 37%, 37% and 48%, respectively for the means, and 44%, 59%, 51%, and 50% for the medians, again showing the difference between poor and affluent schools as a group, but also the impact of a few errant schools.

Using the difference between the school hour flow rate and the midnight flow rate, and the student numbers per school from the WCED dataset, we calculated the effective water used per learner for a six-hour long school day, disregarding the number of educators (and staff), which resulted in 2.17 L/student/day, 1.00 L/student/day, 1.36 L/student/day, and 1.79 L/student/day for the four scenarios, respectively. Since the leaks are already taken into account in these figures, these differences must largely be operational, with the s1 likely higher mainly due to water-intensive feeding schemes, poor maintenance, and lack of user awareness or behaviour; and the s4 number likely higher due to gardening, staff kitchens, and lack of user awareness or behaviour.

These results demonstrate that lack of maintenance at poor schools is a significant contributor to the higher usage at the poorer schools, but also demonstrate that lack of maintenance does not explain the full extent of the differences in usage. We therefore further analysed the underlying drivers.

3.2 Characterising the influence of school affluence on water usage

Prior to quantitative assessment of the interdependencies between water usage and influential factors using BNs modelling, the basic trends between water usage and each influential factor were evaluated. Accordingly, Fig. 2 shows the variations in total water usage of all schools against the number of users, usage during week days, usage during weekends, and usage during different periods of time on week days. It is evident that the total water usage shows strong linear relationships with the usage based on the day and the time of the week. However, there is considerable variability in the relationship between total water usage and the number of students and educators.

Fig. 3 shows the Directed Acyclic Graph *(DAG)* of the two evaluated BNs models incorporating the factors that could influence water usage in schools. These models were fitted with observed data using the *'bnlearn'* package in the R statistical computing platform (Scutari, 2016). Table 3 shows the estimated influence exerted by each factor on school's water usage. Further, Table 3 provides different sets of conditional regression coefficients corresponding to different scenarios of the state of affluence of schools. Additionally, the performance of the proposed model was assessed using leave-one-out cross validation. This resulted a Root Mean Squared Error (RMSE) of 0.0491 and 0.0322 for Model 1 and Model 2, respectively. The observed vs. predicted plots and residuals plots are shown in Fig. 4, which also confirms that the model performance was satisfactory.

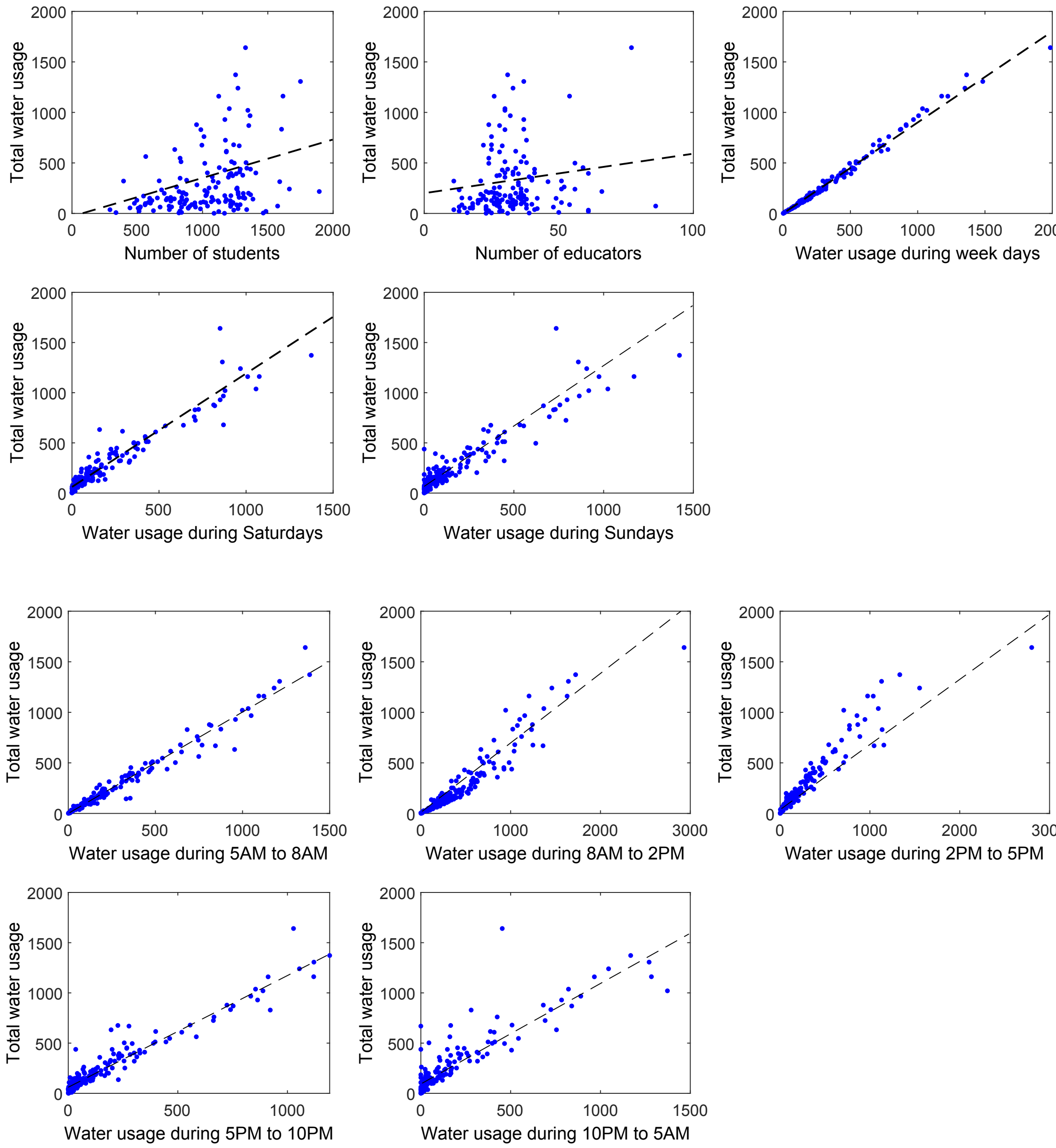

Fig. 2. Total water usage of schools as functions of the number of users, usage during week days, usage during weekend, and usage during days of the week. *Note:* all water usages are

given in L/hour.

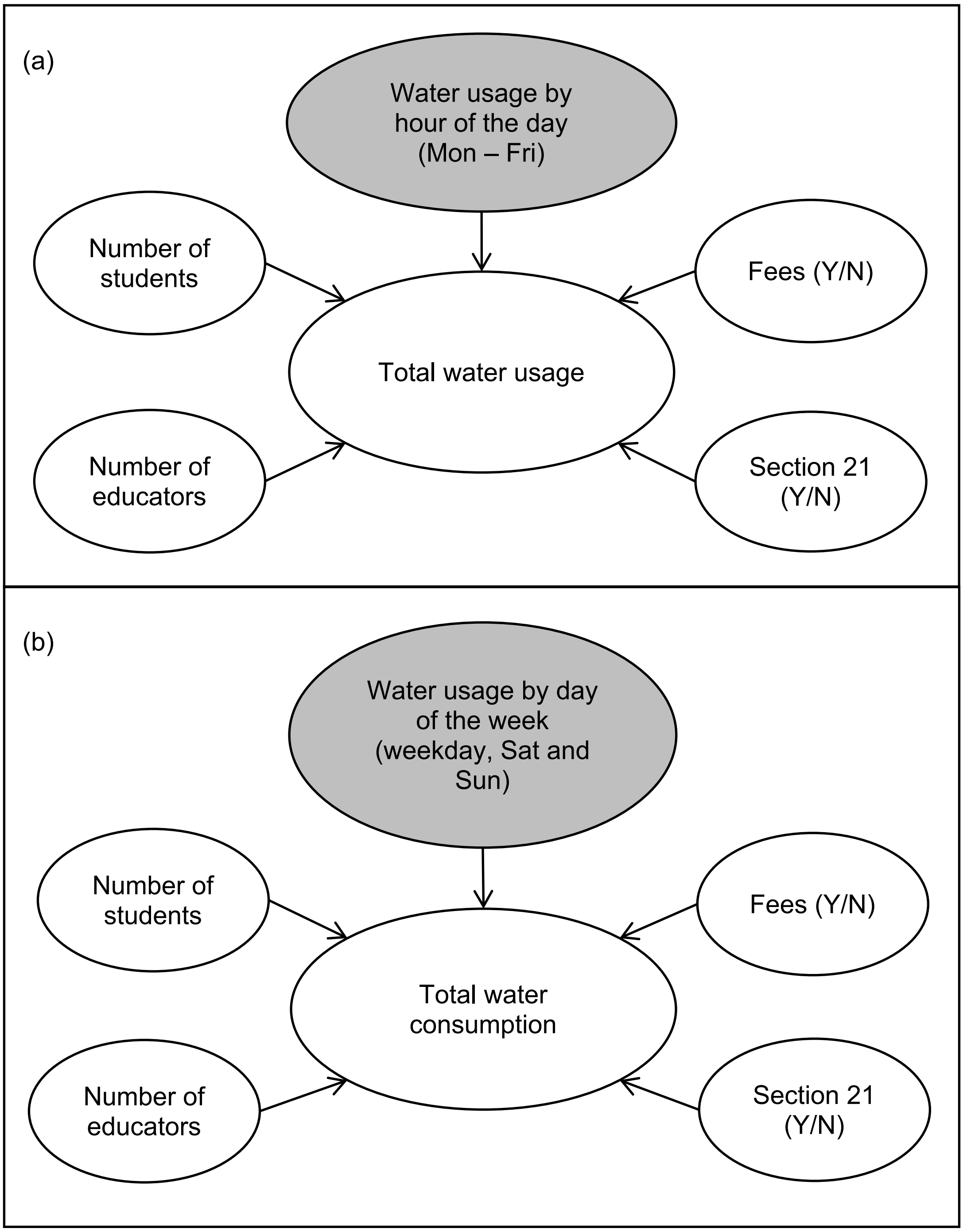

Fig. 3. Directed Acyclic Graph (DAG) of the Bayesian Networks (BNs) models of water usage of schools with different socio-economic status: (a) Model 1; (b) Model 2.

Table 3. Estimated conditional regression coefficients (conditional Gaussian distribution, log transformed data) for total water usage (V_t) and relative influence of key factors for the time-of-the-day usage analysis (Model 1) and day-of-the-week analysis (Model 2).

MODEL 1				
[a]Conditional density: V_t | S21 + Fees + Edu + St + V_{0508} + V_{0814} + V_{1417} + V_{1722} + V_{2205}				
Variable	**Scenario**			
	S1	**S2**	**S3**	**S4**
S21	No	No	Yes	Yes
Fees	No	Yes	No	Yes
	[b]C	**[b]C**	**[b]C**	**[b]C**
Edu	0.028	-0.099	0.034	0.005
St	0.017	0.069	**0.122**	0.034
V_{0508}	**0.353**	**0.381**	**0.588**	**0.270**
V_{0814}	**0.375**	**0.371**	**0.212**	**0.500**
V_{1417}	**0.117**	**0.169**	**0.092**	0.144
V_{1722}	**0.163**	**0.123**	0.046	0.072
V_{2205}	-0.002	0.010	0.004	-0.002
MODEL 2				
[a]Conditional density: V_t | S21 + Fees + Edu + St + V_{sa} + V_{su} + V_w				
Variable	**Scenario**			
	S1	**S2**	**S3**	**S4**
S21	No	No	Yes	Yes
Fees	No	Yes	No	Yes
	[b]C	**[b]C**	**[b]C**	**[b]C**
Edu	0.026	0.147	-0.036	0.035
St	-0.010	-0.184	0.030	-0.040
V_{sa}	**0.148**	**0.115**	0.005	-0.034
V_{su}	0.004	-0.003	0.004	0.042
V_w	**0.855**	**0.911**	**1.017**	**1.006**
[a] probability density function of V_t, given the parent variables				
[b] estimated conditional regression coefficient				

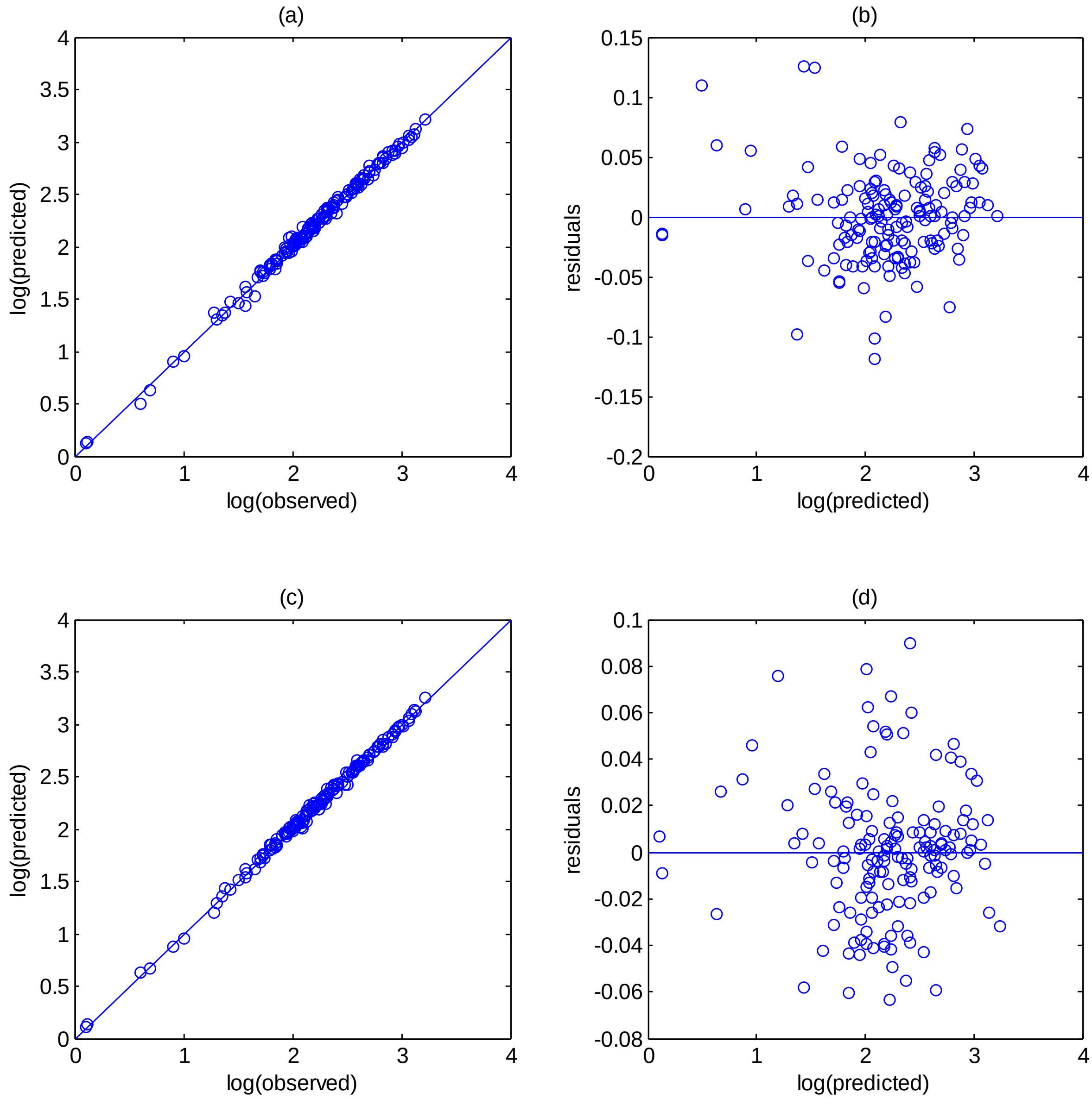

Fig. 4. Predictive performance of the Bayesian Network (BN) models: (a) Model 1 – observed vs predicted plot; (b) Model 1 – residuals plot; (c) Model 2 – observed vs predicted plot; (d) Model 2 – residuals plot.

3.3 Practical implications of research outcomes

The coefficients in Table 3 confirm the hypothesis alluded to in the general assessment. Model 1's coefficients confirm that for the more prudent and affluent Scenario 4, the hours from 05:00 to 17:00 are the main drivers of total usage, with the actual school hours exhibiting by far the largest coefficient.

The coefficients for the schools in the three poorer scenarios show that the hours before school are substantial drivers of total usage. Considering the worst two scenarios (Scenarios 1 and 2), sizeable contributions are evidently made by the hours from 14:00 to 22:00, indicating that it is these that should be targeted to reduce the total usage at schools in these scenarios. The Scenario 3 schools, which perform worse than Scenario 4, but better than the other two has the before-school timeslot as the largest coefficient, and also sizeable contributions from the number of students and the timeslot immediately after school.

Model 2, which evaluated the days of the week, shows that the two worst performing schools have large contributions from the Saturday usage, in addition to the weekday contributions, which are present for Scenarios 3 and 4. Although these results will need further investigation and potentially site inspections, they demonstrate that the total usage at the apparently wasteful schools are largely linked to after hour usage, which is only partially due to constant background leaks.

4.0 Conclusion

We evaluated the temporal usage profiles of 156 schools in the Western Cape in South Africa in the run-up to Cape Town’s Day Zero. We differentiated between affluent and poor schools in general, and specifically evaluated the impact of whether the schools were, fee-paying, self-governed, number of students, number of educators and a diversity of temporal differentiations. The results show a clear trend that the poorest schools (non-fee paying, not self-governing) use substantially more water usage regardless of the time period considered. Moreover, the most affluent schools (fee-paying, self-governing) use the least amount of water. For the two middle schools, we noted usage in-between the two extremities. It was observed that these high levels of usage is likely due to a few errant schools, as the median schools are not much different than the affluent schools. Using Bayesian Networks, it was also observed that in addition to leakages being present, high usage in poorer schools is linked to early morning usage, afternoon and early evening usage as well as Saturday usage. We recommend that schools be equipped with smart meters to allow prudent water usage management. In the alternative, it is recommended that poor schools, especially those that are self-governing be targeted for maintenance and awareness campaigns.

Acknowledgement

The authors thank the Western Cape Department of Education, Shoprite Holdings, Cape Talk, and all the other academic, corporate, and other contributors listed at www.schoolswater.co.za. First and third authors thank MTN South Africa for funding the research through grant S003061, and the Water Research Commission (South Africa) for funding the research through grant K1-7163.

Supplementary Information

A sample of the schools data set of Western Cape Education Department (WCED) and the details of BNs modelling is provided as Supplementary Information.

References

Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B., & Sliusarieva, A. (2012). Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research, 48*(1), W01528 (01521-01514).

Bakker, M., Vreeburg, J., Van Schagen, K., & Rietveld, L. (2013). A fully adaptive forecasting model for short-term drinking water demand. *Environmental Modelling & Software, 48*, 141-151.

Bonotto, D. M., Wijesiri, B., Vergotti, M., da Silveira, E. G., & Goonetilleke, A. (2018). Assessing mercury pollution in Amazon River tributaries using a Bayesian Network approach. *Ecotoxicology and Environmental Safety, 166*, 354-358. doi:https://doi.org/10.1016/j.ecoenv.2018.09.099

Booysen, M. J., Ripunda`, C., & Visser, M. (2019a). Results from a water-saving maintenance campaign at Cape Town schools in the run-up to Day Zero. *Sustainable Cities and Society*. doi:https://doi.org/10.31224/osf.io/45cf9

Booysen, M. J., Visser, M., & Burger, R. (2019b). Temporal case study of household behavioural response to Cape Town's "Day Zero" using smart meter data. *Water Research, 149*, 414-420. doi:https://doi.org/10.1016/j.watres.2018.11.035

Borsuk, M. E., Stow, C. A., & Reckhow, K. H. (2004). A Bayesian Network of Eutrophication Models for Synthesis, Prediction, and Uncertainty Analysis. *Ecological Modelling, 173*(2), 219-239.

Botai, C. M., Botai, J. O., De Wit, J. P., Ncongwane, K. P., & Adeola, A. M. (2017). Drought Characteristics over the Western Cape Province, South Africa. *9*(11), 876.

Bougadis, J., Adamowski, K., & Diduch, R. (2005). Short-term municipal water demand forecasting. *Hydrological Processes: An International Journal, 19*(1), 137-148.

Dalgleish, T., Williams, J. M. G., Golden, A.-M. J., Perkins, N., Barrett, L. F., Barnard, P. J., Au Yeung, C., Murphy, V., Elward, R., & Tchanturia, K. (2007). Reduced specificity of autobiographical memory and depression: the role of executive control. *Journal of Experimental Psychology: General, 136*(1), 23.

Datta, S., Miranda, J. J., Zoratto, L., Calvo-González, O., Darling, M., & Lorenzana, K. (2015). *A behavioral approach to water conservation: evidence from Costa Rica*: The World Bank.

DBE. (2015). *Neims Standard Reports As At 12 May 2015*. Department of Basic Education, Republic of south Africa Retrieved from http://www.education.gov.za/Portals/0/Documents/Publications/NEIMS STANDARD REPORTS AS AT 12 MAY 2015.pdf?ver=2015-06-03-114948-520

Dernoncourt, F., & Lee, J. Y. (2016). *Optimizing neural network hyperparameters with gaussian processes for dialog act classification.* Paper presented at the Spoken Language Technology Workshop (SLT), 2016 IEEE.

Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Alan Roberson, J. (2012). Urban water demand forecasting: review of methods and models. *Journal of Water Resources Planning and Management, 140*(2), 146-159.

Engelbrecht, J., & Harding, A. (2008). The impact of the transition to outcomes-based teaching on university preparedness in mathematics in South Africa. *Mathematics Education Research Journal, 20*(2), 57-70.

Enqvist, J. P., & Ziervogel, G. (2019). Water governance and justice in Cape Town: An overview. *Wiley Interdisciplinary Reviews: Water, e1354*, 1-15. doi:https://doi.org/10.1002/wat2.1354

Ferraro, P. J., & Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics, 95*(1), 64-73.

Fielding, K., Spinks, A., Russell, S., Mankad, A., & Price, J. (2012). Water demand management study: baseline survey of household water use (Part B). *Urban Water Security Research Alliance Technical Report*(93).

Ghiassi, M., Fa'al, F., & Abrishamchi, A. (2017). Large metropolitan water demand forecasting using DAN2, FTDNN, and KNN models: A case study of the city of Tehran, Iran. *Urban Water Journal, 14*(6), 655-659.

Ghiassi, M., Zimbra, D. K., & Saidane, H. (2008). Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning and Management, 134*(2), 138-146.

Government_Gazette. (1996). *South African Schools Act 1996 (Vol. 377, No. 17579)*. Cape Town: Republic of South Africa

Gurung, T. R., Stewart, R. A., Sharma, A. K., & Beal, C. D. (2014). Smart meters for enhanced water supply network modelling and infrastructure planning. *Resources, Conservation and Recycling, 90*, 34-50.

Jacobs, H., & Haarhoff, J. (2004). Application of a residential end-use model for estimating cold and hot water demand, wastewater flow and salinity. *Water sa, 30*(3), 305-316.

Kenney, D. S., Goemans, C., Klein, R., Lowrey, J., & Reidy, K. (2008). Residential water demand management: lessons from Aurora, Colorado1. *JAWRA Journal of the American Water Resources Association, 44*(1), 192-207.

Liu, A., Giurco, D., & Mukheibir, P. (2016). Urban water conservation through customised water and end-use information. *Journal of Cleaner Production, 112*, 3164-3175.

Liu, A., Wijesiri, B., Hong, N., Zhu, P., Egodawatta, P., & Goonetilleke, A. (2018). Understanding re-distribution of road deposited particle-bound pollutants using a Bayesian Network (BN) approach. *Journal of Hazardous Materials, 355*, 56-64. doi:https://doi.org/10.1016/j.jhazmat.2018.05.012

Longueira, R. (2016). *Exploring the functionality of the South African education quintile funding system.* University of Pretoria, Maeda, E. E., Mäntyniemi, S., Despoti, S., Musumeci, C., Vassilopoulou, V., Stergiou, K. I., Giannoulaki, M., Ligas, A., & Kuikka, S. (2017). A Bayesian model of fisheries discards with flexible structure and priors defined by experts. *Ecological Modelling, 366*(Supplement C), 1-14. doi:https://doi.org/10.1016/j.ecolmodel.2017.10.007

Martín de Santa Olalla, F., Dominguez, A., Ortega, F., Artigao, A., & Fabeiro, C. (2007). Bayesian networks in planning a large aquifer in Eastern Mancha, Spain. *Environmental Modelling & Software, 22*(8), 1089-1100. doi:https://doi.org/10.1016/j.envsoft.2006.05.020

McDonald, R. I., Weber, K., Padowski, J., Flörke, M., Schneider, C., Green, P. A., Gleeson, T., Eckman, S., Lehner, B., Balk, D., Boucher, T., Grill, G., & Montgomery, M. (2014). Water on an urban planet: Urbanization and the reach of urban water infrastructure. *Global Environmental Change, 27*, 96-105. doi:https://doi.org/10.1016/j.gloenvcha.2014.04.022

Motala, S. (2015). Equity, access and quality in basic education: a review. *JOURNAL OF EDUCATION*(61), 159-175.

Muller, M. (2018). Lessons from Cape Town's drought. *Nature, 559*, 174-176.

Parks, R., McLaren, M., Toumi, R., & Revett, U. (2019). *Experiences and lessons in managing water from Cape Town*. Retrieved from Grantham Institute, Imerial College, London:

Ren, Z., & Li, S. (2016). *Short-term demand forecasting for distributed water supply networks: A multi-scale approach.* Paper presented at the 2016 12th World Congress on Intelligent Control and Automation (WCICA).

Rigosi, A., Hanson, P., Hamilton, D. P., Hipsey, M., Rusak, J. A., Bois, J., Sparber, K., Chorus, I., Watkinson, A. J., & Qin, B. (2015). Determining the probability of cyanobacterial blooms: the application of Bayesian networks in multiple lake systems. *Ecological Applications, 25*(1), 186-199.

Ripunda, C., & Booysen, M. (2018). *Understanding and affecting school water behaviour using technological interventions.* Paper presented at the Water Institute of Southern Africa (WISA) 2018 Conference, South Africa.

Sánchez-Torija, J. G., Gómez-Rubiera, E. L., & Frutos, C. B. (2017). The incorporation of the study into water consum ption in energy audits in schools. *Revista de la Construcción, 16*(3), 361-373.

Scheba, S., & Millington, N. (2018). Crisis Temporalities: Intersections Between Infrastructure and Inequality in the Cape Town Water Crisis. *International Journal of Urban & Regional Research*.

Scutari, M. (2009). Learning Bayesian Networks with the bnlearn R Package. *arXiv preprint arXiv:0908.3817*.

Scutari, M. (2016). *Package 'bnlearn': Bayesian Network Structure Learning, Parameter Learning and Inference*. Retrieved from

Srinivasan, V., Konar, M., & Sivapalan, M. (2017). A dynamic framework for water security. *Water Security, 1*, 12-20. doi:https://doi.org/10.1016/j.wasec.2017.03.001

Swartz, L. (2009). *Financial Management of Schools*: OpenStax CNX.

Ticehurst, J. L., Newham, L. T. H., Rissik, D., Letcher, R. A., & Jakeman, A. J. (2007). A Bayesian network approach for assessing the sustainability of coastal lakes in New South Wales, Australia. *Environmental Modelling & Software, 22*(8), 1129-1139. doi:https://doi.org/10.1016/j.envsoft.2006.03.003

Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., Rao, P. S. C., Basu, N. B., & Wilson, J. S. (2010). The future of hydrology: An evolving science for a changing world. *Water Resources Research, 46*(5).

WCG. (2017). The Cape Water Crisis – FAQs and honest answers. Retrieved from https://www.westerncape.gov.za/news/water-crisis

Wijesiri, B., Deilami, K., McGree, J., & Goonetilleke, A. (2018). Use of surrogate indicators for the evaluation of potential health risks due to poor urban water quality: A Bayesian Network approach. *Environmental Pollution, 233*, 655-661. doi:https://doi.org/10.1016/j.envpol.2017.10.076

Ziervogel, G. (2019). *Unpacking the Cape Town drought: Lessons learned*. Retrieved from National Treasury, Republic of South Africa: https://www.africancentreforcities.net/wp-content/uploads/2019/02/Ziervogel-2019-Lessons-from-Cape-Town-Drought_A.pdf

A Sample of Western Cape Education Department (WCED) schools data set

NatEmis	Institution_Name	GIS_Longitude	GIS_Latitude	Suburb	Section21	Quintile	Fees?	NAS	tration	NoFeeSchool	Allocation	Urban_Rural	Open_Boarding_school	Full_Service_School	Learners_2016	Educators_2016
100000029	GAIA WALDORF SCHOOL	18.489704	-33.940894	PINELANDS	NO	4	TRUE	N	99	NO	0	URBAN	NO	NO	188	11
100000030	DARUN-NA'IM GIRLS HS	18.470817	-34.012345	WYNBERG	NO		TRUE	N	99	NO	0	URBAN	NO	NO	76	20
100000031	ATLANTIC BEACH COLLEGE	18.43906	-33.7296	MELBOSSTRAND	NO		TRUE	N	99	NO	0	URBAN	NO	NO	30	5
100000036	KNYSNA CHRISTIAN MISSION SCHOOL	23.071752	-34.044101	KNYSNA	NO		TRUE	N	99	NO	0	URBAN	NO	NO	65	5
100000037	EAGLE'S NEST CHRISTIAN PRIM.	18.743455	-33.853628	WALLACEDENE	NO	2	TRUE	N	99	NO	0	URBAN	NO	NO	388	14
100000038	SILVERMINE ACADEMY	18.3984	-34.121929	SUN VALLEY	NO		TRUE	N	99	NO	0	URBAN	NO	NO	84	12
100000054	ACADEMY PRIVATE SCHOOL	18.708624	-33.928723	KUILS RIVER	NO		TRUE	N	99	NO	0	RURAL	NO	NO	54	12
100000055	CLAREMONT HS	18.469091	-33.988005	CLAREMONT	YES	5	TRUE	N	99	NO	193	URBAN	NO	NO	478	26
100000056	WESTLAKE PRIM	18.436522	-34.074136	WESTLAKE	NO	4	FALSE	N	99	YES	1116	URBAN	NO	NO	663	16
100000065	REDDAM HOUSE ATLANTIC SEABOARD	18.408376	-33.910456	GREEN POINT	NO	5	TRUE	N	99	NO	0	URBAN	NO	NO	594	51
100000070	INTERNATIONAL SCHOOL OF HELDERBERG	18.857012	-34.100218	SOMERSET WEST	NO	3	TRUE	N	99	NO	0	URBAN	NO	NO	142	21
100000078	DISA PRIM. (HOUT BAY)	18.361272	-34.027565	HOUT BAY	NO	3	FALSE	N	99	YES	1116	URBAN	NO	NO	537	33
100000091	CEDERBERG ACADEMY	19.014739	-32.585574	CITRUSDAL	NO	4	TRUE	N	99	NO	559	RURAL	NO	NO	725	21
100000103	ECOLE FRANCOISE LE VAILLANT	18.418461	-33.931708	GARDENS	NO	5	TRUE	N	99	NO	0	RURAL	NO	NO	268	37
100000108	NORTHPINE TECHNICAL HS	18.708212	-33.870924	NORTHPINE	NO	5	TRUE	N	99	NO	193	URBAN	NO	NO	298	18
100000109	DELFT TECHNICAL HS	18.629656	-33.965186	DELFT	NO	4	TRUE	N	99	NO	645	URBAN	NO	NO	752	25
100000110	MELKBOS HIGH SCHOOL	18.444036	-33.73001	MELKBOSSTRAND	NO	5	TRUE	N	99	NO	380	URBAN	NO	NO	646	34
100000122	FISANTEKRAAL HS	18.721786	-33.779554	DURBANVILLE	NO	1	FALSE	N	99	YES	1116	URBAN	NO	NO	870	28
100000123	BEAUFORT WEST PRIM	22.569978	-32.351063	BEAUFORT-WEST	YES	3	FALSE	N	99	YES	1116	RURAL	NO	NO	1009	28
100000124	LOUWVILLE HS	18.006815	-32.916368	VREDENBURG	NO	4	FALSE	N	99	YES	1116	RURAL	NO	NO	956	27

Bayesian Networks (BNs) modelling

In BNs modelling, the model structure is developed based on the current knowledge of the system/process being modelled. Therefore, the evidence from the literature and expert opinion play a key role in this regard. The BNs model structure is a Directed Acyclic Graph (DAG), and it is created by connecting a set of random variables which collectively define the system/process of interest (Fig. S1). Then, *Structure Learning Algorithms* are used to learn the model structure, and model parameters are estimated using approaches such as *Maximum Likelihood Estimates* (Ben-Gal, 2007; Scutari, 2009; Uusitalo, 2007).

As depicted in Fig. S1, it is assumed that a particular system/process is represented by a set of random variables $U = \{X_1, X_2, \ldots, X_6\}$. The BNs model structure defines a factorisation of the global probability distribution of U (i.e. joint probability distribution) into local probability distributions of individual variables, based on the *Markov Property* of BNs (Equations 1 and 2). The *Markov Property* states that a particular variable is dependant only on its immediate parent variables (Korb and Nicholson, 2010).

$$P(X_1, X_2, \ldots, X_6) = \prod_{i=1}^{6} P(X_i | \prod X_i) \quad \text{; for discrete variables} \quad (1)$$

$$f(X_1, X_2, \ldots, X_6) = \prod_{i=1}^{6} f(X_i | \prod X_i) \quad \text{; for continuous variables} \quad (2)$$

For discrete variables, the model parameters are estimated in terms of conditional probabilities. For continuous variables, the parameters are estimated in terms of conditional regression coefficients. It is important to note that BNs is a flexible modelling approach, such that a proposed model structure can be improved by using new data and knowledge.

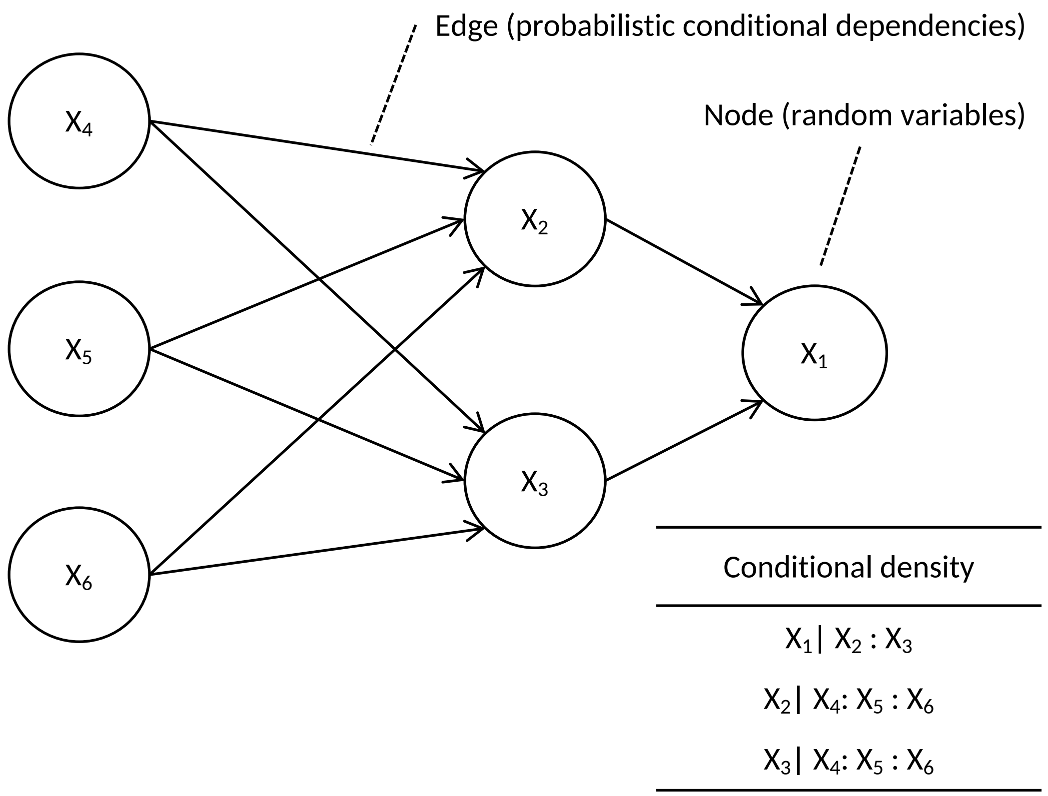

Fig. S1. The graphical structure of a typical Bayesian Networks (BNs) model. *Note:* Conditional density refers to the probability density functions of the variables X_1, X_2 and X_3 given immediate parent variables.

References

Ben-Gal, I. (2007). Bayesian networks. *Encyclopedia of statistics in quality and reliability*.

Korb, K. B., & Nicholson, A. E. (2010). *Bayesian Artificial Intelligence* (2 ed.): CRC press.

Scutari, M. (2009). Learning Bayesian Networks with the bnlearn R Package. *arXiv preprint arXiv:0908.3817*.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling, 203*(3–4), 312-318. doi:http://doi.org/10.1016/j.ecolmodel.2006.11.033

Graphical Abstract