Identification, Prediction, and Explanation of Outdoor Residential Water Consumption Using Smart Meter Data

Valerie Platsko[*1] **and Peter van Beek**[2]

[1,2]Cheriton School of Computer Science, University of Waterloo, Waterloo, ON

* *vplatsko@uwaterloo.ca*

ABSTRACT

Outdoor end-uses of water such as irrigation and filling pools contribute significantly to peak residential consumption. Managing peak consumption in summer is important for reducing infrastructure strain and avoiding potential upgrade costs. We show that outdoor water consumption can be robustly identified from hourly measurement of total water consumption by identifying an upper limit for plausible indoor consumption in a single hour. We also develop regression tree-based models for predicting next-hour water consumption, however the predictability of this consumption is limited. In contrast to previous work, there is little correlation between outdoor consumption and demographic factors such as income. Outdoor consumption shows a large amount of individual variability, with 8.6% of households accounting for 50% of the total outdoor usage.

Keywords: smart meters, peak water demand, irrigation

1 INTRODUCTION

Residential water demand accounts for a significant part of the treated water supplied by municipal water utilities[1], and consumption for outdoor purposes such as irrigation contributes significantly to variability in total water consumption during summer, when demand is highest. We focus on single-family residential water consumption in Abbotsford, British Columbia, where there are concerns about how outdoor consumption contributes to peak demand (the maximum amount of water the system could need to deliver at one time). In addition to ongoing demand management efforts including restrictions on automatic lawn irrigation in summer, Abbotsford installed smart water meters in 2010, in part to help target water conservation messages[2]. Typical smart meters record household water consumption at a higher frequency than the monthly or bi-monthly frequencies required for billing, and measurements every 15 minutes or every hour are common[3]. This increased temporal resolution allows better demand management by showing more precisely when water is consumed, which is important for managing peak demand[4].

Our research has 3 components: identifying outdoor consumption at a household-level from hourly measurements of total consumption, predicting next-hour outdoor and total water consumption for small neighbourhoods, and explaining outdoor water consumption. Typical end-uses of outdoor water (such as lawn irrigation and pool filling) are high volume compared to indoor end uses such as toilet flushing or running taps, making complicated disaggregation techniques requiring high-frequency data less necessary[5], and using hourly data has the advantage that it does not require the installation of additional equipment. Our method of identifying outdoor consumption involves identifying an hourly threshold past which water is likely to be used for mainly outdoor purposes. In contrast to previous work[7], potential alternative thresholds are evaluated using qualitative and quantitative evidence to determine the ideal threshold. Accurate prediction of outdoor consumption would be useful for short-term management of water systems in that it could be used to better estimate the effects of weather conditions and watering restrictions. We develop models for predicting hourly outdoor and total water consumption at a neighbourhood level, although similarly to previous work for prediction at a household level[6], we find predictive accuracy at this scale is limited. These models are the first models for predicting outdoor water consumption at a fine temporal scale (hourly) and a fine spatial scale (small neighbourhoods). Additionally, we analyse the factors which explain outdoor water consumption in Abbotsford during the summer. While the data shows a moderate influence of weather conditions on outdoor consumption, there is little impact from demographic and household variables found to be significant in previous work for other cities, such as income, lot size, and rates of pool ownership.

2 METHODS AND RESULTS

2.1 Datasets

The primary dataset consists of hourly water consumption measurements from the city of Abbotsford, recorded from September 2012 to August 2013. The dataset includes water consumption measurements for 8229 single-family residential units. Secondary datasets were used for demographic and property information. Demographic data was taken from the National Household Survey[1] (NHS). It contains demographic information such as income information and average family size. The NHS data is only publicly available at the census dissemination area (DA) level. (A dissemination area is a contiguous geographic area consisting of multiple census blocks, typically containing 400 to 700 people.) Information about property values and household characteristics was provided by BCAssesment[2]. This information is available at the household level, and contains property values, lot sizes, and building characteristics such as number of bedrooms, and whether the household has a pool.

The water consumption data was aggregated to small neighbourhoods for predicting water consumption, and for explaining the determinants of consumption at a neighbourhood level. These neighbourhoods were chosen as DAs to match the NHS data. The data spans 158 DAs, each containing from 1 to 178 consumers. Because small neighbourhoods retain much of the unpredictability associated with individual consumption, for prediction only, the DAs with fewer than 50 single-family households were omitted. The resulting set is 77 dissemination areas containing 6789 households.

2.2 Identifying Outdoor Consumption

We adapt an approach by Cole and Stewart[7] that makes use of the hourly data already available in Abbotsford. The approach relies on the fact that outdoor consumption (which tends to be high-volume) will cause atypically high hourly volumes of consumption at a household level, and involves developing an hourly threshold such that past this threshold most consumption is for outdoor purposes. Formally, our approach consists of identifying outdoor water consumption (per household,

[1]http://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/prof/index.cfm
[2]https://www.bcassessment.ca/

per hour) as $y_{outdoor} = \max(y_{total} - t, 0)$, and indoor consumption as the remaining consumption, $y_{indoor} = \min(y_{total}, t)$, where t is the identified consumption threshold and y_{total} is the total hourly consumption for the household.

We identify a threshold of 300 L as the threshold t for outdoor consumption. Table 1 contains typical indoor hourly volumes; note that even combinations of indoor uses (such as taking a shower and flushing a toilet in the same hour) are unlikely to reach a threshold of 300 L. Additionally, there are seasonal patterns in water consumption which the threshold should reflect. This threshold is consistent with differences in typical consumption volumes in July and August compared to those in December and January. Figure 1 shows the consumption in summer and winter by hourly volume range. By the 300L range, there is considerably more consumption in summer than in winter.

Table 1: Volumes of typical indoor end-uses, adapted from [8]

End Use	**Volume**
Faucet	4.9 L/min
Toilet	9.8 L/flush
Shower	62 L/shower
Clothes Washer	117 L/load
Dishwasher	23 L/load

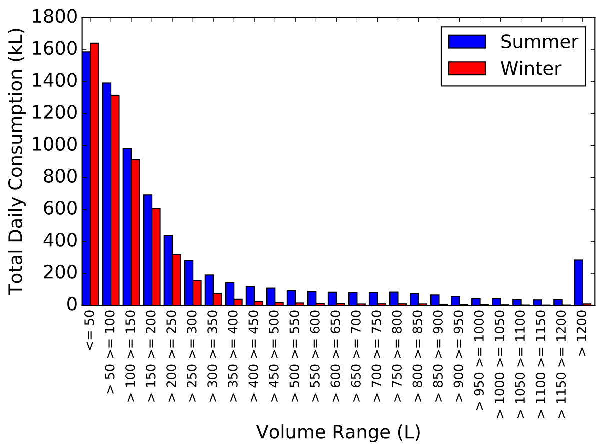
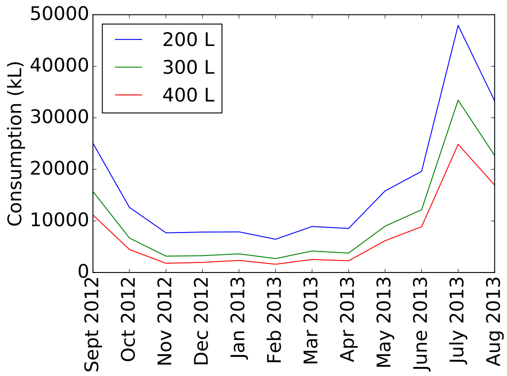


Figure 1: Percentage of consumption by hourly volume, summer and winter

We test the robustness of the threshold approach by evaluating at 3 different thresholds (200 L, 300 L, and 400 L), to determine whether each provides a good estimate given reasonable assumptions about patterns of water consumption. Figure 2 shows the seasonal patterns of indoor and outdoor consumption for each threshold, with clear seasonal patterns for outdoor consumption and only slight seasonal patterns for indoor consumption at all thresholds. Choosing a threshold amounts to choosing a trade-off between identifying all outdoor consumption and not misidentifying any indoor consumption, but our results show the approach is robust. We find indoor consumption at any threshold does not significantly change with household size (estimated based on the number of bedrooms), while outdoor consumption does, which is consistent with typical patterns of consumption[9]. Finally, all thresholds are consistent with the timing of watering restrictions, identifying considerably more outdoor consumption at times when automatic irrigation is permitted.



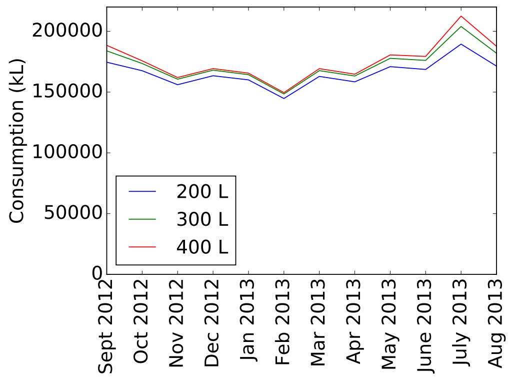


Figure 2: Estimated outdoor (left) and indoor (right) consumption for all households by month, using alternative thresholds

2.3 Predicting Outdoor Water Consumption

Predicting water consumption can be treated as a supervised machine learning task, where the goal is to predict next-hour consumption based on previous consumption volumes and other features. Formally, we fit a model to estimate $\hat{y}_t = \hat{f}(y_{t-1}, y_{t-2}, y_{t-3}, y_{t-168}, x_1, ..., x_n)$, where $\hat{y}_t$ is the predicted value for next-hour water consumption, $\hat{f}$ is the model for prediction, $y_{t-1}, y_{t-2}, y_{t-3}, y_{t-168}$ are hourly values of actual water consumption for the previous 3 hours and the same hour the previous week and $x_1, ..., x_n$ are other features. Water consumption values for the previous several hours, as well as the same hour one week previous, were found to be good predictors of hourly water consumption in previous work[10, 6]. The additional features $x_1, ..., x_n$ include weather information (such as temperature and rainfall), date features indicating weekends and days when watering is permitted, and demographic and property features such as lot size and household income.

Prediction is carried out at the DA level, and the water consumption is normalized by the number of households. $Outdoor$ is our model for predicting outdoor consumption. For comparison, we produce a model $Total$ which predicts total consumption. We fit models based on ensembles of regression trees (using the LSBoost algorithm[11]) to the supervised learning problem described above. This regression tree-based model was chosen because it allows some degree of interpretability through a variable importance measure[12], which essentially shows how much each feature improves predictive accuracy.

The models show a general ability to predict the direction of consumption, although they perform less well at predicting the magnitude of consumption. Figure 3 shows the predicted outdoor and total consumption for the largest dissemination area.. Although overall performance is better for total consumption, the model still fails to predict the magnitude of the morning and afternoon consumption peaks accurately, and sharp peaks in consumption are typically not predicted accurately.

Table 2 shows the absolute errors in litres for each model. Note that for both types of consumption, the average errors are significantly higher than the median, suggesting that the poor performance is in large part due to large errors rather than many small errors. Weekends and watering days when automatic lawn irrigation is permitted have high errors for both models, and both models show errors at similar times of day, suggesting variability in outdoor consumption also makes predicting total consumption difficult.

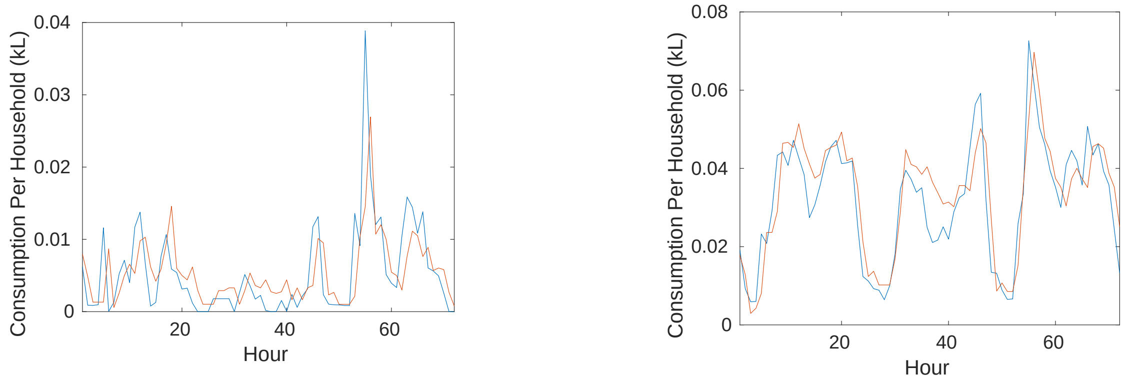


Figure 3: Prediction of Outdoor (left) and Total (right) compared to actual value for the largest dissemination area in the first 3 days of July. The orange line is the predicted consumption value and the blue line is the actual value.

Table 2: Absolute errors (litres) for Outdoor and Total, showing average and 50th, 75th, and 90th percentile values

	Outdoor				***Total***			
	average	**50th**	**75th**	**90th**	**average**	**50th**	**75th**	**90th**
overall	5.054	2.819	6.453	12.122	11.725	8.648	16.259	25.842
weekdays	4.877	2.752	6.281	11.783	11.495	8.497	15.999	25.361
weekends	5.558	3.027	7.010	13.297	12.380	9.074	17.059	27.258
watering	5.407	3.040	6.890	12.813	12.113	8.871	16.722	26.670
no watering	4.612	2.572	5.938	11.338	11.240	8.377	15.649	24.861

While weather and demographic features are included in the model, the most significant predictive features are previous values of water consumption. Figure 4 shows the variable importances for each model. In both cases, the previous hour's consumption is the most relevant variable. The second most relevant variable for every model is y_{t-168}. Both y_{t-1} and y_{t-168} are excluded from the figures in order to better show the importances of the other variables. The results suggest previous consumption is more predictive at this timescale than demographic or weather-related factors.

2.4 Determinants of Outdoor Water Consumption

Explainable variation in outdoor water consumption between households or between neighbourhoods can be used for demand management and planning watering restrictions[7]. We find that, aside from pool ownership, there is little explainable variation caused by the factors that we studied, however, there is considerable individual variability between households. Our statistical results use Pearson's correlation coefficient (r) to measure linear relationships, and Spearman's rank correlation coefficient (ρ) to show the strength of a monotonic relationship. The analysis is restricted to summer (July and August) when water demand is highest.

By ordering the households by their consumption, it is possible to show the percentage of water consumed by the highest-consumption households. Figure 5 shows the cumulative consumption for the first n households ordered by outdoor consumption. 25% of outdoor water is used by the 174

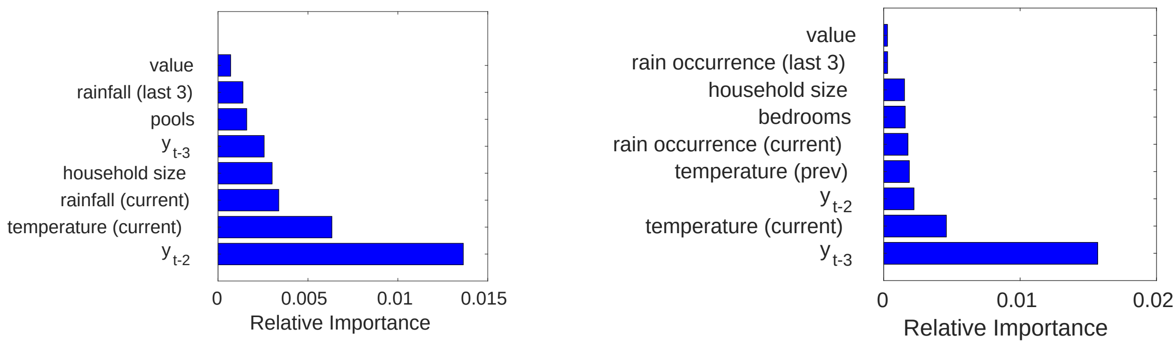


Figure 4: Variable importances for Outdoor (left) and Total (right), with y_{t-1} and y_{t-168} omitted

households with the highest consumption. 704 households (8.6%) are responsible for 50% of the outdoor consumption, and just over half of households (4305) consume 95% of the water used for outdoor purposes, with the remaining households using almost insignificant amounts of outdoor water. The contribution of the highest-consumption households to total outdoor usage is an important consideration for demand management and also partially explains the difficulty of predicting outdoor consumption for small neighbourhoods.

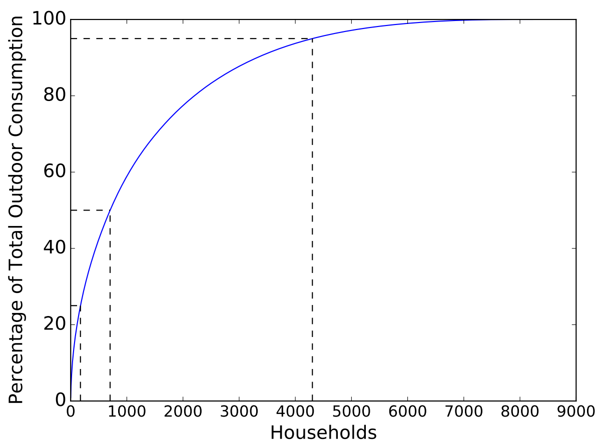


Figure 5: Cumulative consumption for first n households, ordered by amount of outdoor water consumption and showing the number of consumers responsible for 25%, 50% and 95% of consumption

Because a major source of outdoor water demand is irrigation, it is expected that outdoor water consumption will decrease with rainfall. There are clear decreases in outdoor consumption when rainfall occurs, as shown in Figure 6. Despite similar average temperatures (19.4 °C and 19.1 °C), the total outdoor consumption for July 2013 was 1037 kL. compared to 698 kL a day in August 2013. This can be explained by the significantly higher and more frequent rainfall in August (57 mm over 9 days) as compared to in July (1.6 mm on a single day). There also is a weak positive correlation ($\rho = 0.401, r = 0.409$) between temperature and outdoor water consumption during July and August.

The demographic and property factors analysed at a neighbourhood level were income ($\rho = 0.122, r = 0.0939$), lot size ($\rho = 0.0745, r = 0.423$), and rates of pool ownership ($\rho = -0.0396, r = 0.145$). Note that the lack of correlation with rates of pool ownership is not surprising, given that the largest

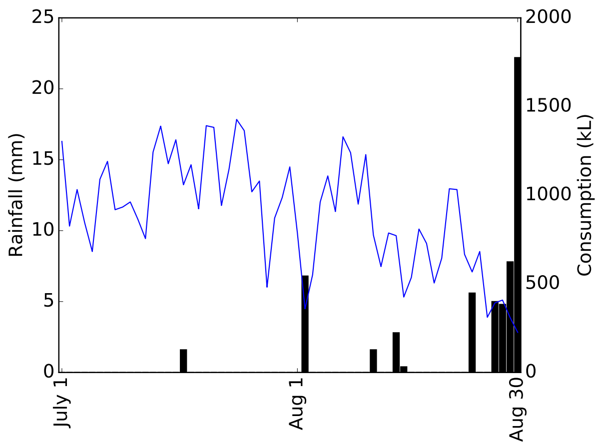


Figure 6: Rainfall and outdoor water consumption for all households in July and August 2013

percentage of pools in a DA is only 9%. While rates of pool ownership do not significantly affect outdoor water consumption at the dissemination area level, there are significant differences in water consumption between individual households with and without pools. As shown in Figure 7, typical water consumption is much higher for households that have pools. The median consumption is 24.7 kL for households with pools and 4.7 kL for households without pools. We find, conversely, little correlation ($\rho = 0.0817, r = 0.144$) with lot size. Because many households do not use significant amounts of outdoor water, the analysis is restricted to the 703 highest-consumption households which account for 50% of outdoor water consumption. The lack of correlation may be because lot size is relatively consistent between households, with almost all households having relatively small lots. Additionally, no information about ground cover or type of landscaping was available, which has been shown to affect water consumption for irrigation[13].

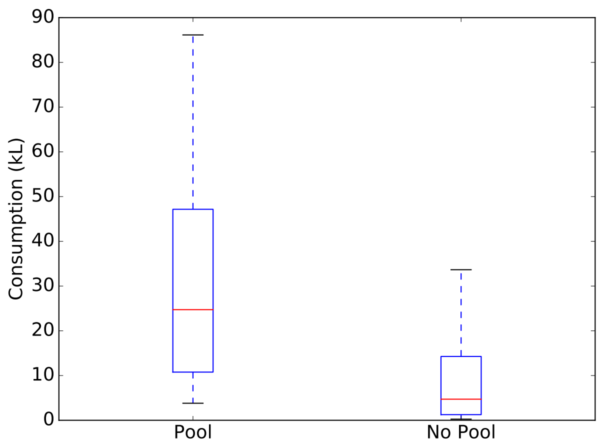


Figure 7: Variability of outdoor water consumptions for households with and without pools, showing the 5th, 25th, 50th, 75th, and 95th percentiles of consumption

3 CONCLUSIONS

Our results suggest outdoor water consumption in Abbotsford can be identified with sufficient accuracy from smart meter readings to partially explain outdoor water consumption and to identify the highest-consuming households. We find significant between-household variability in outdoor water consumption which is not explained by the demographic factors studied. This has important implications for managing peak demand, targeting conservation efforts, and estimating the effects of weather on summer water consumption, but makes predicting outdoor water consumption for small neighbourhoods difficult. Future work could include determining the amount of spatial aggregation required for accurate prediction.

References

[1] Environment Canada, "2011 municipal water use report : Municipal water use 2009 statistics," 2011.

[2] K. Filiatrault and T. Kyle, "Smart meter update," Abbotsford/Misson Water & Sewer Commission, Tech. Rep. ENG 45-2011, 2013.

[3] G. Hauber-Davidson and E. Idris, "Smart water metering," *Water*, vol. 33, no. 3, pp. 38–41, 2006.

[4] T. Boyle, D. Giurco, P. Mukheibir, A. Liu, C. Moy, S. White, and R. Stewart, "Intelligent metering for urban water: A review," *Water*, vol. 5, no. 3, pp. 1052–1081, 2013.

[5] D. Carboni, A. Gluhak, J. A. McCann, and T. H. Beach, "Contextualising water use in residential settings: A survey of non-intrusive techniques and approaches," *Sensors*, vol. 16, no. 5, p. 738, 2016.

[6] D. Walker, E. Creaco, L. Vamvakeridou-Lyroudia, R. Farmani, Z. Kapelan, and D. Savić, "Forecasting domestic water consumption from smart meter readings using statistical methods and artificial neural networks," *Procedia Engineering*, vol. 119, pp. 1419–1428, 2015.

[7] G. Cole and R. Stewart, "Smart meter enabled disaggregation of urban peak water demand: Precursor to effective urban water planning," *Urban Water Journal*, vol. 10, no. 3, pp. 174–194, 2013.

[8] W. B. DeOreo, P. W. Mayer, B. Dziegielewski, and J. Kiefer, "Residential end uses of water, version 2," 2016.

[9] P. W. Mayer, W. B. DeOreo, E. M. Opitz, J. C. Kiefer, W. Y. Davis, B. Dziegielewski, and J. O. Nelson, "Residential end uses of water," 1999.

[10] M. Herrera, L. Torgo, J. Izquierdo, and R. Pérez-García, "Predictive models for forecasting hourly urban water demand," *Journal of Hydrology*, vol. 387, no. 1, pp. 141–150, 2010.

[11] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, pp. 1189–1232, 2001.

[12] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*. Springer series in statistics New York, 2001, vol. 1.

[13] B. Breyer, H. Chang, and G. Parandvash, "Land-use, temperature, and single-family residential water use patterns in Portland, Oregon and Phoenix, Arizona," *Applied Geography*, vol. 35, no. 1, pp. 142–151, 2012.