An advective-dispersive transport model for residential water consumption

Robert W. Enouy, Andre J. A. Unger and Rashid Rehan

ABSTRACT

This work applies an advective-dispersive framework to simulate utility-wide residential water consumption using the analogy of a continuum transport process. In this context, the advective-dispersive process describes how changes in real water price and seasonal weather variability influence water consumption distribution, which ultimately governs mean and total water consumption values. Water consumption response is measured using histogram data optimally fit using parametric probability density functions (PDF) that have consistent parametrization over the entire observation period. Median statistic denotes advection and prescribes location of the measurement-space PDF, while standard deviation combined with standard-score PDF denotes dispersion which provides the measurement-space PDF with scale and shape. Combining location, scale, and shape components produces a measurement-space PDF that represents the solution to advective-dispersive transport phenomena. We use a Taylor series expansion of the statistics that define the PDF along with curvilinear regression to develop constitutive relationships that define how location, scale, and shape of the PDF respond to price and weather information. This results in a fully parametrizing advective-dispersive process represented by a partial differential equation that provides a tool for anticipating the probability that households will experience water poverty or use excess amounts as price, weather, and policy considerations change through time. This approach is conducive to automation when combined with smart water metering.

Key words | advection, dispersion, policy, price, water consumption, weather

Robert W. Enouy
Andre J. A. Unger (corresponding author)
Department of Earth and Environmental Sciences,
University of Waterloo,
Ontario,
Canada
E-mail: *andre.unger@uwaterloo.ca*

Rashid Rehan
National Institute of Urban Infrastructure Planning,
University of Engineering and Technology,
Peshawar,
Pakistan

INTRODUCTION

Water utilities need to accurately forecast residential water consumption so that they can adjust the unit price of water and balance revenues with expenses, enabling them to operate their system on a full-cost recovery basis. United States Environmental Protection Agency (EPA) documents a shift in policy goals over the past 20 years toward the financially sustainable operation and management of water utilities (EPA 2003, 2005, 2006). In Ontario, Canada, the Water Opportunities and Conservation Act mandates that municipal water utilities develop sustainability plans for their water distribution services (MOE 2011).

Hughes & Leurig (2013) suggest that changing water consumption habits have resulted in considerable revenue uncertainty, potentially sabotaging utility efforts to develop financially sustainable management practices. House-Peters & Chang (2011) and Donkor *et al.* (2011) provide a comprehensive review of the advances in methodologies for urban water forecasting and analysis that quantify trends in water consumption, such as econometrics, agent-based, system dynamics, and artificial neural-network models. They conclude that the increased data richness has led to improved modeling techniques; however, they

suggest that future work will need to develop novel techniques to incorporate this information and ultimately elucidate water consumption relationships at multiple scales. Furthermore, they identify three key characteristics that may lead to a generally accepted water consumption modeling framework:

- Development of water consumption models for practical application should focus on those models with input variables that can be easily collected, monitored, and used by the utility.
- Water consumption models should be as parsimonious as possible without compromising the integrity of their forecasting quality.
- Future development should focus on probabilistic forecasting methods that allow utilities to make decisions, while quantifying the level of uncertainty of the resulting water consumption forecasts.

These authors acknowledge that there is no clear answer to the question 'Which model is best for water consumption forecasting?' and state that current water consumption modeling applications require specific parameterization and implementation for different geographic locations, water rate structures, historical data quality, and periodicity. Therefore, there is a need in this industry to develop a general water consumption forecasting approach that broadly applies by overcoming the specificity of the current models and methodologies.

The objective of this analysis is to identify and quantify ambient processes that represent external stresses which correlate with temporal shifts in the collective residential water demand distribution when aggregated into a probability mass function (PMF) (i.e., a histogram). We hypothesize that a probabilistic analysis tool for forecasting water consumption will benefit the financial sustainability of water utilities. This analysis tool uses a continuum analogy to study how consumers collectively adjust their water consumption behavior to these external stresses. In this study, external stresses include tangible and hence measurable factors such as real water price, temperature, and precipitation. However, additional stresses may include less tangible factors such as water conservation initiatives, education, and by-law enforcement. Hereafter, we refer to these external stresses as 'ambient processes' under the

pretext that we first must determine processes that are causal to water consumption behavior given that all residential water consumption accounts are equally exposed to these external stresses. The continuum analogy is predicated on the observation that these residential water demand PMFs can be accurately replicated as continuous probability density functions (PDFs). Moreover, the functional dependence of statistics that describe location, scale, and shape of these PDFs can then be parameterized via a statistically significant regression to these external stresses. We then present a mathematical framework that transforms the relationship between the statistics that describe location, scale, and shape of these PDFs as a function of environmental stresses into an 'advective-dispersive'-like transport model for the transient evolution of the probability density of water consumption. This probability density is quantified by solving a partial differential equation with respect to time as well as the external stresses. A key outcome of this approach is to be able to anticipate changes to targeted populations including water impoverished households and excessive consumption within the utility. Combining the approach with smart metering technology could promote automation within water utilities and assist in understanding the influence of policy, pricing, and demographic changes on residential water demand and hence revenue.

The foundation of this analysis is the set of residential water consumption data collected from the City of Waterloo, Ontario, Canada. Specifically, these data comprise water consumption meter readings representing ten years consisting of 60 bimonthly billing periods between January/February 2007 and November/December 2016 for a total of 1,549,371 observations and 51,291,348 m^3 of cumulative billed water. The pricing structure of the water utility is a volume-constant rate and the utility services upwards of 27,000 residential accounts during each billing period. Figure 1 provides PMF data from the November/December bimonthly billing period for the years 2008, 2010, 2012, and 2014. PMF data is computed using a frequency histogram that counts the number of consumers in 1 m^3 water consumption bins and subsequently divides this number by the total measurements within a sampling interval. Hence, the area under the PMF is unity. The utility has been annually increasing its real water price, with nearly a one-real dollar per cubic meter (40%) increase between
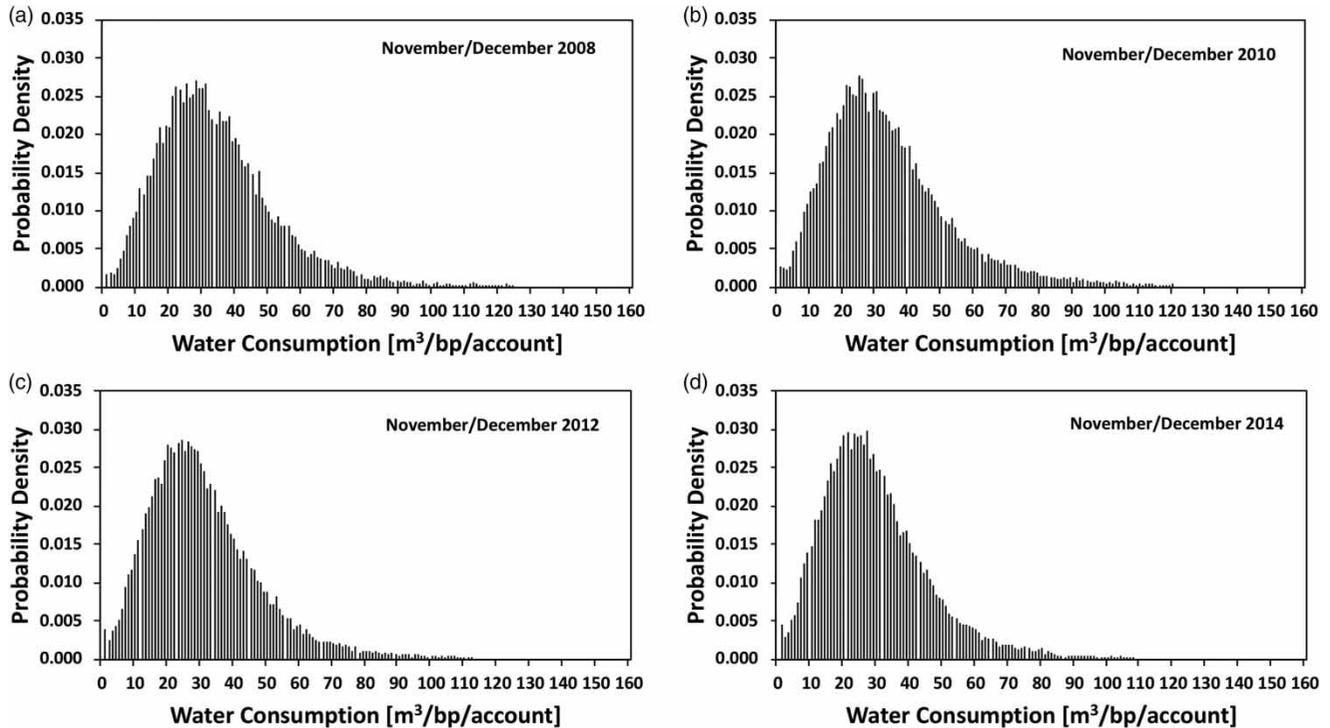
Figure 1 │ Select residential water consumption PMFs for November/December from the City of Waterloo, Ontario, Canada. Note that weather conditions are generally consistent during the November/December billing periods.

2008 and 2014. This annual price increase qualitatively results in declining water consumption characterized by a progressive compression of the water consumption PMF toward the origin.

The methodology begins by transforming all of the discrete histograms for each of the 60 bimonthly billing periods into smooth and continuously differentiable PDFs. This step involves measuring discrete statistics for each sampling period, and then choosing a representative but normalized parametric PDF p_z in the standard-score space which has a consistent form throughout the analysis. All sampling periods are characterized by a unimodal PDF that is shifted and asymmetric with a heavy tail. The data are denoted to exist in the space x which in the case of water consumption has units of $[m^3/billing\ period/account]$. The median of the dataset m_x is used to measure the location of the water consumption PMF, the standard deviation σ_x measures its scale, while the control function parametrization for the standard-score PDF p_z characterizes its shape. The standard score variable z is defined as $z = \dfrac{x - m_x}{\sigma_x}$. This methodology closely resembles that of

kernel density estimation techniques for econometrics applications (Zambom & Dias 2012). However, the consistent parametrization of the control function is critical to assess how the shape of the water consumption PDF continuously evolves through time as a function of ambient processes.

The outline of this paper proceeds as follows. First, we present the discrete statistics that measure the location, scale and shape of the best-fit PDF to each water consumption PMF. Second, we develop the mathematical framework to correlate the functional parameterization of each of the location, scale, and shape statistics to the transient external stresses. This framework is based on Taylor series leading to a multivariate curvilinear regression. This regression provides the basis for testing the validity of our continuum analogy. Third, we apply this framework to the City of Waterloo residential water consumption dataset. Validation of the framework follows by ensuring that the arithmetic mean of the data used to construct each PMF is identical to within measurement error of the mean statistic of its best fit PDF. The impact of omitted external stresses with particular emphasis on water conservation policy is

discussed in the context of failure of the advective-dispersive transport model to correctly replicate the water consumption PMF during times of policy enforcement. Fourth, we demonstrate the usefulness of the advective-dispersive transport framework to calculate the probability that consumers exist in either excessive or impoverished water consumption ranges. Previous publications have identified this information as key for policy intervention to protect the most vulnerable and target conservation efforts to heavy water users (McGranahan 2002; Gargano *et al.* 2017).

METHODS

Advective-dispersive transport is used to model the temporal evolution of utility-wide residential water consumption as a function of changes in ambient processes, such as the unit price of water and weather conditions. The water consumption response at any observation interval can be represented by a histogram, which is then transformed into a PMF and finally a PDF which allows for calculation of the mean statistic. Consequently, the temporal evolution of the PDF and mean statistic represents the solution to an advective-dispersive transport problem for a continuum system. Here, we use the analogy of a continuum system to represent utility-wide residential water consumption. The location, scale, and shape of a measurement space PDF p_x are quantified using the median m_x, standard deviation σ_x, and standard-score PDF p_z, respectively. The relationship between the PDF p_x and the control function g_z is summarized in Appendix A (available with the online version of this paper). In this context, the median represents the bulk translation (advection) of the distribution, while the standard deviation and standard-score PDF combine to characterize the relative frequency or spread (dispersion) of the data as:

$$\underbrace{p_x}_{\substack{continuum \\ distribution}} = \underbrace{m_x}_{\substack{probabilistic \\ advective \\ process}} + \underbrace{\frac{1}{\sigma_x} \times p_z}_{\substack{probabilistic \\ dispersive \\ process}} \tag{1}$$

A summary derivation for Equation (1) is provided in Appendix A. Appendix A also demonstrates that the mean

statistic of the measurement space data μ_x is a scalar value that describes the ensemble magnitude of the measurement space PDF p_x as a continuum system. Using this premise, the mean statistic also represents an advective-dispersive process as:

$$\mu_x = \underbrace{m_x}_{\substack{distribution \\ location}} + \underbrace{\sigma_x}_{\substack{distribution \\ scale}} \times \underbrace{\mu_z}_{\substack{distribution \\ symmetry}} \tag{2}$$
$$\underbrace{}_{\substack{mean \\ advective \\ process}} \qquad \underbrace{}_{\substack{mean \\ dispersive \\ process}}$$

where the standard-score mean μ_z quantifies the symmetry of the distribution and has a value of $\mu_z = 0$ for symmetric distributions. The symmetry of the distribution is solely dependent upon the definite integral of the position-weighted standard-score PDF p_z and the control function parameters that describe it. Given the conditional dependence of μ_z on p_z in the context of advective-dispersive transport in Equation (2), we can present the mean statistic as a projection of the PDF symmetry through its location and scale as: $\mu_z = \int z p_z dz$. Notably, a distribution has perfect symmetry for values of $\mu_z = 0$ with this value growing as the distribution becomes increasingly asymmetric. Perfect symmetry results in the advective-dispersive process of the mean in Equation (2) being solely dependent upon the median m_x.

Development of the advection-dispersion model for p_x and μ_x in the sections below begins by evaluating the discrete statistics of the raw data from each sampling period and then transforming them into a time-continuous form. This results in a time-continuous PDF whose parametric values can be adjusted so as to be able to reproduce the entire sampling sequence of discrete histogram information. The parametric values allow the PDF to change location, scale, and shape in response to a set of continuum ambient processes, which change as a function of time t over sequential sampling intervals. These ambient processes are denoted using the variable 'x_j'. Note that the following model development remains general with respect to the relationship between the statistics for the median m_x, standard deviation σ_x, as well as the standard-score PDF p_z and the ambient processes $x_j = f(t)$.

Discrete statistics

The discrete statistics are scalar measures of the location $m_{x,i,t}$, scale $\sigma_{x,i,t}$, and mean statistic $\mu_{x,i,t}$ of a 'continuum system' because they reflect the aggregation of all of the active accounts within a sampling period t into a single distribution. Notably, t represents the sampling interval of the analysis described by each discrete statistic. The arithmetic mean characterizes the magnitude of the discrete data for some sampling interval t and can be expressed as follows:

$$\mu_{x,i,t} = \frac{1}{N_{i,t}} \sum_{i=1}^{N_i} x_{i,t}, \ \mu_{z,i,t} = \frac{\mu_{x,i,t} - m_{x,i,t}}{\sigma_{x,i,t}} \tag{3}$$

where water consumption measured for any account i in the discrete sampling interval (billing period) t is denoted as $x_{i,t}$ and $N_{i,t}$ represents the number of active residential accounts within the utility at each sampling interval. Knowledge of the discrete median $m_{x,i,t}$ and standard deviation $\sigma_{x,i,t}$ allow for transformation of the mean $\mu_{x,i,t}$ into a representation of distribution symmetry $\mu_{z,i,t}$ in the standard-score space. The discrete median $m_{x,i,t}$ and standard deviation $\sigma_{x,i,t}$ are quantified for each billing period as:

$$m_{x,i,t} = \begin{cases} Y_{(N_{i,t}+1)/2}, & \text{if } N_{i,t} \text{ is odd} \\ \frac{1}{2}(Y_{N_{i,t}/2} + Y_{1+N_{i,t}/2}), & \text{if } N_{i,t} \text{ is even} \end{cases}$$

$$\sigma_{x,i,t} = \sqrt{\frac{1}{(N_{i,t}-1)} \sum_{i=1}^{N_{i,t}} [x_{i,t} - m_{x,i,t}]^2} \tag{4}$$

where Y_n represents the order of the set of water consumption measurements $x_{i,t}$ from minimum value Y_1 to the maximum value $Y_{N_{i,t}}$ (Hogg & Craig 1995). Finally, the discrete median $m_{x,i,t}$ and standard deviation $\sigma_{x,i,t}$ relate to the statistical transformations from the measurement space x as: median-relative space $y = \frac{x}{m_{x,i,t}}$; and, standard-score space $z = \frac{x - m_{x,i,t}}{\sigma_{x,i,t}}$. Pearson (1894) first introduced the standard deviation of the dataset x_i relative to the arithmetic mean $\mu_{x,i}$. Equation (4) combines the median absolute deviation introduced by Gauss (1816) with the idea of squaring the deviation and is equivalent to Pearson's interpretation of standard deviation for symmetric distributions where $m_{x,i,t} = \mu_{x,i,t}$. Equation (2) demonstrates

that the mean statistic $\mu_{x,t}$ is a function of the median, standard deviation, and the PDF $p_{z,t}$. In this spirit, evaluating the standard deviation as a function of the median prevents a recursive relationship between the standard deviation and mean value for asymmetric datasets.

Upon evaluating the discrete statistics for each sampling interval, the analysis can hypothesize a formulation of $p_{z,t}$ for reproducing the shape of the PMF at sampling interval t as:

$$p_{z,t} = \exp\left(-\int g_{z,t} dz\right),$$
$$g_{z,t} = [\alpha_{1,t} + \alpha_{2,t}z + \alpha_{3,t}z^2 + \alpha_{4,t}z^3] \tag{5}$$

where $p_{z,t}$ represents the continuous standard-score PDF; $g_{z,t}$ is its third-order exponential polynomial control function; and $\alpha_{n_z,t}$ represents the control function parameters for each histogram at sampling interval t. Notably, the standard-score PDF $p_{z,t}$ for each sampling interval transforms into the measurement space PDF $p_{x,t}$ using the discrete median $m_{x,i,t}$ and standard deviation $\sigma_{x,i,t}$ using Equation (1) as: $p_{x,t} = m_{x,i,t} + \frac{1}{\sigma_{x,i,t}} p_{z,t}$.

The probability weighted mean $\mu_{z,t}$ or symmetry of the distribution in the standard score space z is derived from the parametric PDFs $p_{z,t}$ at each sampling interval t as:

$$\mu_{z,t} = \int_{z_{min}}^{z_{max}} z p_{z,t} dz \tag{6}$$

where $\quad z_{min} = \frac{x_{min} - m_{x,i,t}}{\sigma_{x,i,t}} \quad$ and $\quad x_{min} = 0;$ and $z_{max} = \frac{x_{max} - m_{x,i,t}}{\sigma_{x,i,t}}$ and $x_{max} = 4m_{x,i,t}$ from data culling (note: we return to this issue of data culling when discussing the residential water consumption application later in this work). Equation (6) shows a relationship between symmetry $\mu_{z,t}$ and the control function $g_{z,t}$ through the standard-score PDF $p_{z,t}$ and is the discrete representation of the symmetry estimator for data within each sampling interval. Evaluating the discrete standard-score mean $\mu_{z,t}$ for the corresponding PDF during each sampling interval t requires numerical integration of Equation (6). Finally, $\mu_{z,t}$ can be transformed from the standard-score space z into the measurement space x as $\mu_{x,t}$ using the discrete median

$m_{x,i,t}$ and standard deviation $\sigma_{x,i,t}$ from Equation (2) as:

$$\mu_{x,t} = m_{x,i,t} + \sigma_{x,i,t}\,\mu_{z,t}.$$

Time-continuous statistics

Time-continuous statistics are an extension of the above discrete statistics in that they are a continuous function of an ambient process as well as time. To begin the process of defining the continuous statistics, the total derivative of each statistic is expanded with respect to two independent variables x_1 and x_2 that represent the relevant ambient processes of interest. Begin by defining a continuous statistic as $U \in \{\mu_x, m_x, \sigma_x, \alpha_{n_z}\}$ and $U = f(x_1, x_2)$:

$$dU = F'_U dx_1 + F^*_U dx_2 \tag{7}$$

where $\alpha_{n_z} \in \{\alpha_1 \dots \alpha_4\}$ for a third-order control function. F'_U represents the partial derivative of U with respect to x_1 as $F'_U = \dfrac{\partial U}{\partial x_1}$ where $\dfrac{\partial U}{\partial x_1} = f(x_1, x_2)$. Similarly, F^*_U represents the partial derivative of U with respect to x_2 as $F^*_U = \dfrac{\partial U}{\partial x_2}$ where $\dfrac{\partial U}{\partial x_2} = f(x_1, x_2)$. Consider that $x_1 = f(t)$ and $x_2 = f(t)$ are time-dependent processes. Then it follows that the statistics describing a PDF can inherit their time-dependence from these processes: $U = f(x_1, x_2, t)$. Therefore, with knowledge of the time-ordered nature of x_1 and x_2, the influence of these ambient processes can be projected onto the progression of the statistics that describe the distribution and its mean. This results in an advective-dispersive representation of a PDF as it and its mean statistic evolve through time from Equation (3). Equation (7) hypothesizes a model that x_1 and x_2 will correlate with the continuum statistics that describe the discrete histogram. What remains is a formal expansion of F'_U and F^*_U for each statistic with respect to these independent variables. Taking the definite integral of the total derivative from Equation (7) from the position at point $x_1 = 0$ and $x_2 = 0$ results in a general representation of the function $U = f(x_1, x_2, t)$ as:

$$dU = \int\limits_{x_1=0}^{x_1} F'_U dx_1 + \int\limits_{x_2=0}^{x_2} F^*_U dx_2 \tag{8}$$

Next, expand F'_U and F^*_U using a multivariate Taylor expansion around the point $x_1 = 0$ and $x_2 = 0$ as:

$$
\begin{aligned}
F'_U &= \int\limits_{x_2=0}^{x_2} \frac{\partial F'_U}{\partial x_2} dx_2 + \int\limits_{x_1=0}^{x_1}\left[\int\limits_{x_2=0}^{x_2} \frac{\partial^2 F'_U}{\partial x_1 \partial x_2} dx_2\right] dx_1 \\
&\quad + \int\limits_{x_1=0}^{x_1}\left[\int\limits_{x_1=0}^{x_1}\left[\int\limits_{x_2=0}^{x_2} \frac{\partial^3 F'_U}{\partial x_1^2 \partial x_2} dx_2\right] dx_1\right] dx_1 + \dots \\
F^*_U &= \int\limits_{x_1=0}^{x_1} \frac{\partial F^*_U}{\partial x_1} dx_1 + \int\limits_{x_2=0}^{x_2}\left[\int\limits_{x_1=0}^{x_1} \frac{\partial^2 F^*_U}{\partial x_1 \partial x_2} dx_1\right] dx_2 \\
&\quad + \int\limits_{x_2=0}^{x_2}\left[\int\limits_{x_2=0}^{x_2}\left[\int\limits_{x_1=0}^{x_1} \frac{\partial^3 F^*_U}{\partial x_2^2 \partial x_1} dx_1\right] dx_2\right] dx_2 + \dots
\end{aligned}
\tag{9}
$$

Then, substitute and integrate F'_U and F^*_U within the total derivative to generate a curvilinear regression model for $U = f(x_1, x_2, t)$. To compress notation let $F''_U = \dfrac{\partial F'_U}{\partial x_1}$, $F'''_U = \dfrac{\partial^2 F'_U}{\partial x_1^2}$; $F^{**}_U = \dfrac{\partial F^*_U}{\partial x_2}$, $F^{***}_U = \dfrac{\partial^2 F^*_U}{\partial x_2^2}$; $F^{*\prime}_U = \dfrac{\partial F'_U}{\partial x_2}$, $F^{*\prime\prime}_U = \dfrac{\partial^2 F'_U}{\partial x_1 \partial x_2}$; $F^{*\prime}_U = \dfrac{\partial F^*_U}{\partial x_1}$, $F^{**\prime}_U = \dfrac{\partial^2 F^*_U}{\partial x_1 \partial x_2}$; and, $F^{*\prime\prime\prime}_U = \dfrac{\partial^3 F'_U}{\partial x_1^2 \partial x_2}$ and $F^{***\prime}_U = \dfrac{\partial^3 F^*_U}{\partial x_2^2 \partial x_1}$. Substitution into Equation (9) produces a general relationship that solves for $U = f(x_1, x_2, t)$ as:

$$
\begin{aligned}
dU &= \int\limits_{x_1=0}^{x_1}\left[\int\limits_{x_2=0}^{x_2} F^{*\prime}_U dx_2 + \int\limits_{x_1=0}^{x_1}\left[\int\limits_{x_2=0}^{x_2} F^{*\prime\prime}_U dx_2\right] dx_1\right. \\
&\quad \left. + \int\limits_{x_1=0}^{x_1}\left[\int\limits_{x_1=0}^{x_1}\left[\int\limits_{x_2=0}^{x_2} F^{*\prime\prime\prime}_U dx_2\right] dx_1\right] dx_1 + \dots\right] dx_1 \\
&\quad + \int\limits_{x_2=0}^{x_2}\left[\int\limits_{x_1=0}^{x_1} F^{*\prime}_U dx_1 + \int\limits_{x_2=0}^{x_2}\left[\int\limits_{x_1=0}^{x_1} F^{**\prime}_U dx_1\right] dx_2\right. \\
&\quad \left. + \int\limits_{x_2=0}^{x_2}\left[\int\limits_{x_2=0}^{x_2}\left[\int\limits_{x_1=0}^{x_1} F^{***\prime}_U dx_1\right] dx_2\right] dx_2 + \dots\right] dx_2
\end{aligned}
\tag{10}
$$

This expression for dU can be used to determine how the mean statistic μ_x changes as a function of x_1 and x_2. Moreover, dU can be adapted into the form of a transport

model to estimate the median m_x, standard deviation σ_x, and control function parameters $\alpha_{n_z,\ell}$. The intent of the transport model is to reproduce the trends of the entire PDF and indirectly evaluate the mean statistic as a function of ambient processes $\mathbb{x}_1$ and $\mathbb{x}_2$.

Advective-dispersive transport with ambient processes

To proceed with adapting Equation (10) for curvilinear regression, first truncate the multi-variate Taylor series expansion at $F_U^{***\prime} = F_U^{*\prime\prime\prime} = 0$ so that $dU \cong d\hat{U}$.

$$dU = F'_{U,0}\mathbb{x}_1 + F^*_{U,0}\mathbb{x}_2 + 2F^{*\prime}_U\mathbb{x}_2\mathbb{x}_1 + \frac{1}{2}F^{**}_{U,0}\mathbb{x}_2^2$$
$$+ \frac{1}{2}F^{**\prime}_U\mathbb{x}_2^2\mathbb{x}_1 + \frac{1}{2}F''_{U,0}\mathbb{x}_1^2 + \frac{1}{2}F^{*\prime\prime}_U\mathbb{x}_2\mathbb{x}_1^2 \tag{11}$$

where $F'_{U,0}$ is the constant of integration obtained from evaluating the definite integral $\int_{\mathbb{x}_2=0}^{\mathbb{x}_2} F^{*\prime}_U \, d\mathbb{x}_2$; $F^*_{U,0}$ is the constant of integration obtained from $\int_{\mathbb{x}_1=0}^{\mathbb{x}_1} F^{*\prime}_U \, d\mathbb{x}_1$; and similarly, the constants of integration $F''_{U,0}$ and $F^{**}_{U,0}$ result from their corresponding integrals in Equation (10). Next, all partial derivatives in Equation (11) are expressed as coefficients $b_{U,1...7}$ to succinctly express $\hat{U}$ as a curvilinear regression model:

$$\hat{U} = b_{U,0} + b_{U,1}\mathbb{x}_1 + b_{U,2}\mathbb{x}_2 + b_{U,3}\mathbb{x}_2\mathbb{x}_1 + b_{U,4}\mathbb{x}_2^2$$
$$+ b_{U,5}\mathbb{x}_2^2\mathbb{x}_1 + b_{U,6}\mathbb{x}_1^2 + b_{U,7}\mathbb{x}_2\mathbb{x}_1^2 \tag{12}$$

where $\int d\hat{U} = \hat{U} - b_{U,0}$ and $b_{U,0...7}$ represent the curvilinear regression parameters.

To proceed with a general representation of Equation (1) as an advective-dispersive transport model under the influence of ambient processes, first condense the notation in Equation (12) with $\hat{U} \in \{\hat{m}_x, \hat{\sigma}_x, \hat{\alpha}_{1...4}\}$ and $\widehat{U} = f(b_{U,0...7}, \mathbb{x}_1, \mathbb{x}_2, t)$. Next, re-express Equation (1) as:

$$\hat{p}_x = \hat{m}_x + \frac{1}{\hat{\sigma}_x}\exp\left(-\int[\hat{\alpha}_1 + \hat{\alpha}_2 z + \hat{\alpha}_3 z^2 + \hat{\alpha}_4 z^3]dz\right) \tag{13}$$

Similarly, advective-dispersive transport of the mean statistic following Equation (2) is expressed as:

$$\hat{\mu}_{x|\hat{p}_x} = \hat{m}_x$$
$$+ \hat{\sigma}_x \int_{z_{min}}^{z_{max}} z \exp\left(-\int[\hat{\alpha}_1 + \hat{\alpha}_2 z + \hat{\alpha}_3 z^2 + \hat{\alpha}_4 z^3]dz\right) dz \tag{14}$$

where z_{min} and z_{max} is the range of the integration in accordance with Equation (6). Both Equations (13) and (14) indicate that advective-dispersive transport occurs as the location, scale, and shape of the continuum distribution of observed measurements respond to ambient processes.

Validation of the advective-dispersive transport process for the mean statistic follows by using Equation (12) to directly regress its response to ambient processes. Notation for this process is given as:

$$\hat{\mu}_x = f(b_{\mu_x,0...7}, \mathbb{x}_1, \mathbb{x}_2, t) \tag{15}$$

Equation (14) is referred to as the 'transport model mean' or $\hat{\mu}_{x|\hat{p}_x}$, while Equation (15) is referred to as the 'direct regression mean' or $\hat{\mu}_x$. Equivalence of (14) and (15) implies that the magnitude of the ensemble continuum response to ambient processes can be inferred without knowledge of the location, scale, and shape of the distribution of observations itself. However, this information obviously exists and serves to constrain the range of measurable data constituting the continuum response, as represented by the PMF. Moreover, the transport model $\hat{p}_x$ guarantees a unique solution for the mean statistic $\hat{\mu}_{x|\hat{p}_x}$ by replicating the PMF of the raw data. This is in contrast to the direct regression mean $\hat{\mu}_x$ which does not constrain combinations of the location, scale, and shape of the distribution. Therefore, working with the direct regression mean $\hat{\mu}_x$ ignores the availability of the location, scale, and shape information describing the PMF and hence the relative probability that a residential account will consume a particular amount of water.

RESULTS AND DISCUSSION

This analysis uses a continuum analogy to understand how aggregate residential water consumers respond to changes in price, weather, and policy. Figures A1 and A2 in Appendix A

(available with the online version of this paper) support this analogy by independently showing a shift in the location, scale, and shape of the best-fit PDF to price and seasonal weather patterns for each measurement period. Note that residential water consumption occurs continuously throughout a billing period and does not represent a discrete one-time usage. Moreover, price and weather patterns also change continuously (even as residential water consumption for any given account is instantaneously adjusted at each point in time) throughout a billing period as climate patterns, economic conditions, or municipal policies evolve. Therefore, we wish to remind the reader that discrete measurements of water consumption, weather, and water price for a specific sampling interval are indicative of a time-continuous process.

The previous section develops the transport model to quantify the continuum response of residential water consumption as changes to the PDF $\hat{p}_x$ with respect to external stresses. In addition to real water price P, it is anticipated that weather score W as a representation of temperature and precipitation, water restriction by-law enforcement, water conservation, and education are key features impacting water consumption. While P and W are observable ambient processes that can be measured and recorded, public policy initiatives such as by-law enforcement and education are difficult to quantify in the same manner. However, the response of the water consumption histogram to changes in P, W is tangible. It is expected that the impact of policy and education on water consumption can only be inferred via observed changes in the water consumption PDF beyond those that can be explained via tangible ambient processes. The development of the continuous statistics above, culminating in Equation (13), shows that the advective-dispersive transport of $\hat{p}_x$ is dependent on knowledge of the median, standard deviation, and control function statistics of the water consumption dataset. Discrete statistics $m_{x,i,\ell}$, $\sigma_{x,i,\ell}$, and $\mu_{x,i,\ell}$ for each of the sampling 60 bimonthly periods ℓ are itemized in Table B1 in Appendix B (available with the online version of this paper). Here, this analysis denotes the set of these continuum ambient processes that change during each sampling period ℓ using the variable $\mathbb{x}_j \in \{P(\ell), W(\ell)\}$.

This analysis proceeds in three steps. First, it applies the methodology from Enouy (2018) to transform water consumption histograms (see Figure 1) into optimally parameterized and continuous PDFs that are consistent with the advective-dispersive processes expressed in Equations (1) and (2). Second, the analysis performs curvilinear regression upon the median, standard deviation, and control function parameters with real water price and weather score. Statistically defensible correlation to ambient processes supports the contention that the water consumption PMF represents a continuum system that experiences advective-dispersive transport. Third, the analysis compares the direct regression model to the transport model estimates for the mean statistic. This section builds experimental evidence using the above advection-dispersive transport theory to support the continuum analogy and infer that water consumption PMFs exhibit a continuum response to ambient processes. The outcome of the following analysis is that the transport model is at least as effective as the direct regression model for estimating the evolution of the mean statistic while also providing information on the probability $\mathcal{P}$ of residential water consumption occurring within any prescribed interval $\Delta x = x_{max} - x_{min}$ where $\mathcal{P} = \int_{x_{min}}^{x_{max}} p_x(x, P, W, \ell)dx$ for any billing period ℓ.

Parametric PDFs as a representation of the continuum response

Enouy (2018) describes the methodology for estimating the control function parameters in order to optimally fit the PMF data shown in Figure 1. The control function produces a 'best fit' parametric PDF $p_{z,\ell}$ representing a continuous function that compresses each dataset at sampling period ℓ into a median $m_{x,i,\ell}$, standard deviation $\sigma_{x,i,\ell}$, and control function parameters $\alpha_{1,\ell}$, $\alpha_{2,\ell}$, $\alpha_{3,\ell}$, and $\alpha_{4,\ell}$. The discrete statistics $m_{x,i,\ell}$ and $\sigma_{x,i,\ell}$ provide scalar estimates of the location and scale of the observation data $x_{i,\ell}$ while the shape of the cumulative mass function (CMF) is captured by $p_{z,\ell}$ (see Equation (5)) through the control function $g_{z,\ell}$. These statistics are estimated by matching cumulative distribution functions (CDFs) to the CMFs derived from the culled data. The resulting control function parameters and CMFs for each bimonthly period ℓ between January/February 2007 through November/December 2016 are itemized in Table B2 in Appendix B as well as Tables B3(a) to B3(j), respectively. As previously mentioned, this parametrization proceeds on the basis that limited data culling of the observation data $x_{i,\ell}$ is necessary to remove
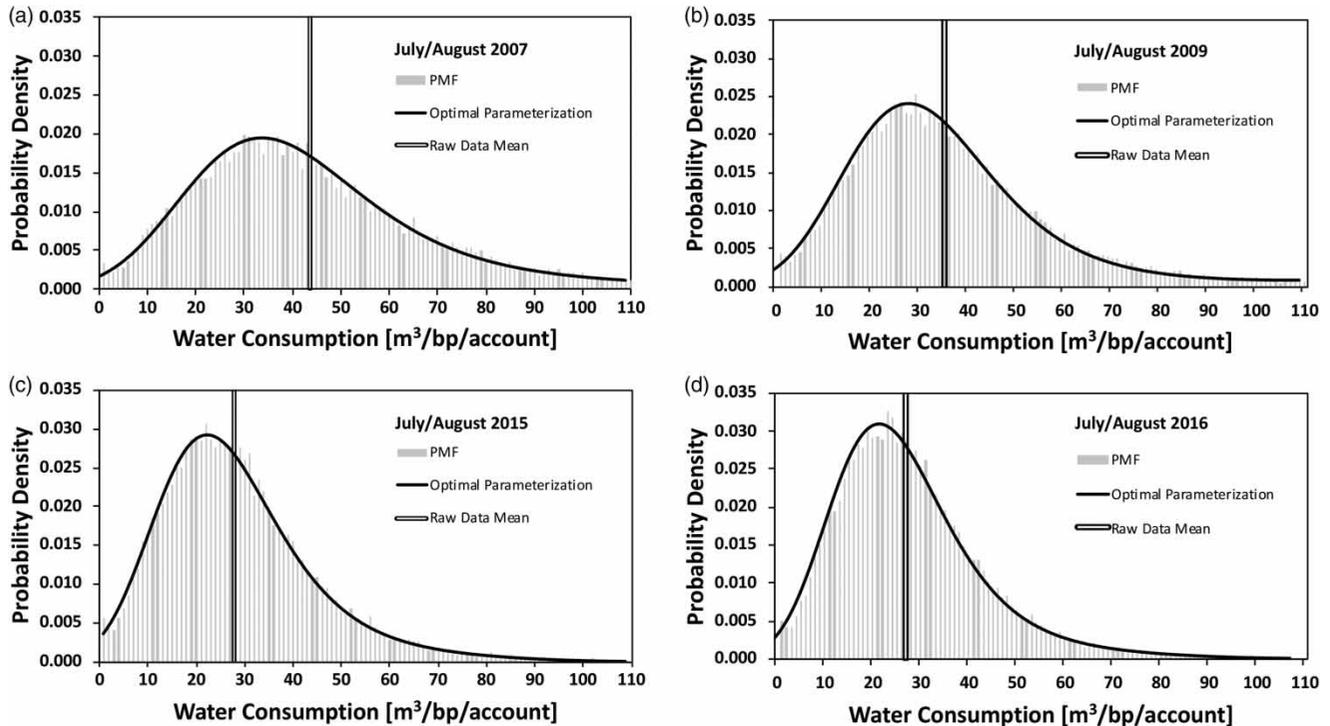
Figure 2 | Select water consumption data and optimal parametric fit. Residential water consumption PMFs and their corresponding best fit PDFs from July/August billing period. Also shown is the arithmetic mean $\mu_{x,i,\ell}$ of the raw data.

measurements that are greater than four times the median $m_{x,i,\ell}$ of each sampling period ℓ: in other words, data for which $y_{i,\ell} > 4$ are removed. The remaining data reflect greater than 98% of the original data for all billing periods considered herein. Water consumption measurements beyond this threshold include multi-unit dwellings and extreme residential water consumers, which do not reflect the water consumption behavior for the population of interest in this analysis – single-family dwellings.

The control function parameterization is adjusted to ensure that the shape of the CDF matches that of the CMF for each sampling period. This step provides evidence that $p_{x,\ell}$ reproduces the continuum process represented by the PMF for sampling period ℓ, and that $\mu_{x,\ell}$ is a unique representation of $\mu_{x,i,\ell}$. Table A1 in Appendix A presents the mean square error estimates $MSE = \sqrt{[\mu_{x,\ell} - \mu_{x,i,\ell}]^2}$ for each sampling interval to quantify the departure of the continuous distribution from the raw data. MSE values of the mean water consumption are always less than $1 \, m^3/bp/acct$ which is the measurement accuracy for each meter reading $x_{i,\ell}$, where '*bp*' denotes a billing period and

'*acct*' denotes account. Figure 2 presents the PDF with the optimal parameterization for the July/August billing period during 2007, 2009, 2015, and 2016 superimposed onto its respective PMF to demonstrate the goodness of fit over the entire range of observation data $x_{i,\ell}$. To varying degrees, each water consumption PMF is reproduced by PDFs that are asymmetric, shifted, and exhibit a heavy tail.

Price and weather as ambient processes

In the previous section, this analysis supports that the ensemble residential accounts behave as a continuum process replicated by the water consumption PDF $p_{x,\ell}$. The next step is to quantify the correlation and infer causality in terms of how transient ambient processes influence the location $m_{x,i,\ell}$, scale $\sigma_{x,i,\ell}$, and shape parameters ($\alpha_{1,\ell}$, $\alpha_{2,\ell}$, $\alpha_{3,\ell}$, and $\alpha_{4,\ell}$) of the water consumption PDF $p_{x,\ell}$ as well as the corresponding mean statistic μ_x during each sampling period ℓ. Ambient processes can be either macroscopic or microscopic in terms of their influence on consumers. Macroscopic ambient processes are experienced equally by

all consumers within the utility, and include temperature, precipitation, real water price, education, and by-law enforcement. Intuitively, macroscopic processes should drive advection of the water consumption PDF through scaling of the median statistic. Additionally, they could also influence the scale and shape of the PDF provided the population of residential accounts experiences a heterogeneous response to changes in these utility-wide macroscopic processes. Microscopic processes are only experienced by a subset of the population and may include changes in household income and number of occupants. Microscopic processes may not have an influence on a sufficient number of consumers to cause advection of the water consumption PDF. However, changes experienced by a subset of the population could influence dispersion through adjustments to the scale and shape of the PDF.

Price P_t is measured at each sampling period t and represents real water price as the depreciated variable unit cost of metered water. Prices are discounted using the annual consumer price index (CPI) inflation rate to a base year of 2004$. This analysis applies CPI under the assumption that it reflects increases in household income for all residential accounts, hence any price increases above CPI represent real changes in water affordability relative to household income. Appendix C (available with the online version of this paper) shows a strong negative correlation between price and the annual average water consumption. The weather score W_t

at each sampling period t is a function of rainfall R_t and temperature T_t measurements combined into a single process as:

$$W_t = T_t \times R_t \qquad (16)$$

where T_t represents the average of the daily high temperature in degrees Celsius for all days within sampling period t (University of Waterloo Weather Station 2016); and, R_t represents the number of days with less than 2 mm of rainfall during sampling interval t (NASA 2016; Environment Canada 2017).

The weather score is based on the hypothesis that the influence of temperature T_t and rainfall R_t on water consumers are variables that cannot be separated. Based on a correlation analysis presented in Appendix C, this rendition of W_t shows moderate positive correlation with the annual average water consumption. Note that while the utilities cannot control the weather score W_t, they are able to adjust the real water price P_t to ensure revenues generated from the variable unit cost of water promote financial sustainability. However, utilities adjust their water price once per year in advance of unknown seasonal weather variations within the target billing year. This minimizes the interdependence between the utility-controlled water price and seasonal variation in the weather score.

Figure 3 presents the discrete values for weather score and real water price for all billing periods between
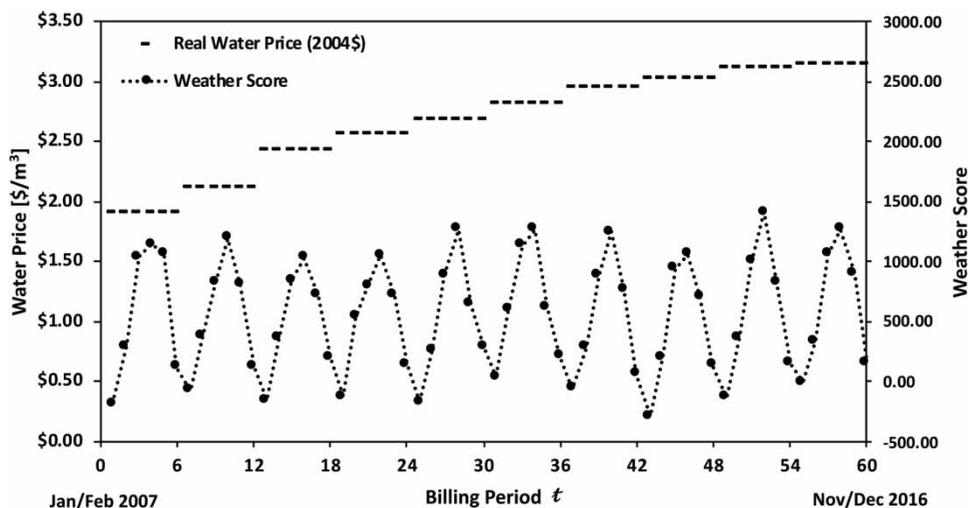

Figure 3 | Ambient conditions for water consumption during entire analysis period. Time series representation of price P_t and weather score W_t variables for each bimonthly sampling period t. See Appendix B, Table B4 (available with the online version of this paper).

January/February 2007 and November/December 2016. The utility annually increases the real water price to boost their revenues, while the weather score changes periodically due to seasonal variability in temperature and precipitation. The troughs that appear along the weather score visualization represent the winter months, whereas the peaks represent summer months. Variability in the amplitude and width of the peaks are a consequence of seasonal weather variability that may include extreme weather events such as heavy rainfall in March/April and May/June or drought conditions in July/August and September/October billing periods. A sensitivity analysis on the influence of the 'dry days' threshold is provided in Appendix C.

Transport model parameterization

Previously, water consumption data from the City of Waterloo were used to establish the idea that the resulting PMF could be replicated with a parametric PDF, and evolution of this PDF represents the continuum response of the utility-wide residential accounts to real water price and weather conditions. The next step is to parametrize the coefficients within Equations (13) and (15) representing advective-dispersive transport of the PDF $\hat{p}_x$ and the mean statistic $\hat{\mu}_x$. Equations (11) and (12) show that these coefficients are partial derivatives of the location $m_{x,i,t}$, scale $\sigma_{x,i,t}$, and shape parameters ($\alpha_{1,t}$, $\alpha_{2,t}$, $\alpha_{3,t}$, and $\alpha_{4,t}$) with respect to real water price $\mathbb{x}_1 \sim P(t)$ and weather score $\mathbb{x}_2 \sim W(t)$. Advective-dispersive transport along with its partial-derivative coefficients are a time-continuous process. This requires $t \to t$ resulting in a water consumption PDF $p_x(x, P, W, t)$ as well as its mean statistic $\mu_x(P, W, t)$ representing a transient one-dimensional process along the axis of water consumption x $[m^3/bp/acct]$.

Multi-variate curvilinear regression is used to estimate model parameters $b_{0...7}$ within Equation (12) for the time-continuous statics $\hat{U} \in \{\mu_x, m_x, \sigma_x, \alpha_{1...4}\}$ where $\hat{U} = f(b_{U,0...7}, P, W, t)$. Estimation of model parameters $b_{0...7}$ parameterizes the advective-dispersive transport model for the water consumption PDF $\hat{p}_x(x, P, W, t)$ and its resulting mean statistic $\hat{\mu}_{x|p_x}(P, W, t)$ given by Equations (13) and (14), respectively. Additionally, these model parameters also result in a description of the direct regression

mean $\hat{\mu}_x(P, W, t)$ given by Equation (15). Table A2 in Appendix A summarizes results from the multi-variate curvilinear regression performed on the full suite of data itemized in Appendix B, Tables B1, B2, and B4. Regression results are for the general form of Equation (12) with the dependence of $\hat{U} \in \{\mu_x, m_x, \sigma_x, \alpha_{1...4}\}$ on $P(t)$ and $W(t)$, with model parameters $b_{U,0...7}$ removed (set to '0.00') when p-values were greater than a 10% significance level.

The curvilinear regression results show that each of the mean $\hat{\mu}_x$, median $\hat{m}_x$, standard deviation $\hat{\sigma}_x$, and control function parameters $\hat{\alpha}_{1...4}$ statistically correlate with the observed changes to P and W. The p-value on the F statistic indicates that there is less than 1% chance that any one relationship is coincidental, with the mean, median, and standard deviation showing stronger correlation than the control function parameters. Contributing parameters $b_{0...7}$ vary between each statistic, which may indicate that the mean value as well as the location, scale, and shape of the distribution are controlled by different processes. Table 1 summarizes the active model parameters for each statistic as well as their derivative representation from the total derivative and Taylor series expansion given by Equations (8) and (9).

The R^2 value for the median statistic is larger than that for either the mean or the standard deviation, which indicates that the variance of the median is reproduced more accurately than for either the mean or standard deviation.

Table 1 │ Active model parameters for each statistic

Parameter		$\hat{\mu}_x$	$\hat{m}_x$	$\hat{\sigma}_x$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$
$b_{U,0}$	*constant*	✓	✓	✓	✓	✓	✓	✓
$b_{U,1}$	$\dfrac{\partial U}{\partial P}$	✓	✓	✓	–	✓	–	–
$b_{U,2}$	$\dfrac{\partial U}{\partial W}$	–	✓	–	✓	–	–	–
$b_{U,3}$	$\dfrac{\partial^2 U}{\partial P \partial W}$	–	✓	–	–	✓	✓	✓
$b_{U,4}$	$\dfrac{\partial^2 U}{\partial W^2}$	✓	✓	✓	–	–	–	–
$b_{U,5}$	$\dfrac{\partial^3 U}{\partial P \partial W^2}$	✓	✓	✓	–	✓	–	–
$b_{U,6}$	$\dfrac{\partial^2 U}{\partial P^2}$	–	–	–	–	✓	–	–
$b_{U,7}$	$\dfrac{\partial^3 U}{\partial P^2 \partial W}$	–	–	–	–	–	✓	–

✓ *active model parameter*; – *inactive model parameter*.

Note that the standard deviation is dependent upon the median statistic (see Equation (4)) and may inherit its estimation error. Contributing coefficients $b_{0...7}$ for the control function shape parameters ($\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, and $\hat{\alpha}_4$) vary and exhibit low R^2 values, which indicates that combinations of W and P only partially reflect observed variability. Three possible explanations exist for the inability of the curvilinear model to more accurately reproduce the observed variability in each statistic. First, unaccounted macroscopic ambient processes as well as microscopic household processes impact water consumption beyond that which can be explained by price and weather alone. These were itemized earlier but may not necessarily be limited to: passive water conservation, education, and by-law enforcement, as well as household income and number of occupants. Second, imprecision of water consumption measurements recorded to within $1\,m^3$ severely restricts accuracy in the discrete median statistic $m_{x,i,t}$. This may influence the sensitivity of the model given that the discrete median $m_{x,i,t}$ is used to estimate the standard deviation $\sigma_{x,i,t}$ and each control function parameter $\alpha_{n_z,t}$, and ultimately influences the transformation between the measurement space x, the median-relative space y, and the standard score space z. This seems to have the greatest impact on estimating parameters for the control function model with their characteristic low R^2 values. Third, using time-averaged weather score data W_t that span bimonthly sampling periods may restrict the sensitivity of the transport model from expressing the influence of severe and localized weather events. It is expected that shorter sampling intervals could provide greater resolution in the water consumption response to extreme seasonal weather conditions.

This advective-dispersive transport model (from Equation (13)) representing the continuum response of the utility-wide residential water demand to the ambient process of real water price and weather score is now shown for the PDF solution $\hat{p}_x(x, P, W, t)$ as:

$$
\hat{p}_x(x, P, W, t)
$$
$$
= \hat{m}_x(P, W, t) + \frac{1}{\hat{\sigma}_x(P, W, t)} \exp\left(-\int [\hat{\alpha}_1(P, W, t) + \hat{\alpha}_2(P, W, t) z \right.
$$
$$
\left. + \hat{\alpha}_3(P, W, t) z^2 + \hat{\alpha}_4(P, W, t) z^3] dz\right) \tag{17}
$$

Similarly, the derived transport mean from Equation (14) $\hat{\mu}_{x|\hat{p}_x}(P, W, t)$ is expressed as:

$$
\hat{\mu}_{x|\hat{p}_x}(P, W, t) = \hat{m}_x(P, W, t)
$$
$$
+ \hat{\sigma}_x(P, W, t) \int_{z_{min}}^{z_{max}} z \exp\left(-\int [\hat{\alpha}_1(P, W, t) + \hat{\alpha}_2(P, W, t) z \right.
$$
$$
\left. + \hat{\alpha}_3(P, W, t) z^2 + \hat{\alpha}_4(P, W, t) z^3] dz\right) dz \tag{18}
$$

where

$$
\hat{m}_x(P, W, t) = b_{m,0} + b_{m,1} P(t) + b_{m,2} W(t)
$$
$$
+ b_{m,3} W(t) P(t) + b_{m,4} W(t)^2 + b_{m,5} W(t)^2 P(t)
$$
$$
\hat{\sigma}_x(P, W, t) = b_{\sigma,0} + b_{\sigma,1} P(t) + b_{\sigma,4} W(t)^2 + b_{\sigma,5} W(t)^2 P(t)
$$
$$
\hat{\alpha}_1(P, W, t) = b_{\alpha_1,0} + b_{\alpha_1,2} W(t)
$$
$$
\hat{\alpha}_2(P, W, t) = b_{\alpha_2,0} + b_{\alpha_2,1} P(t) + b_{\alpha_2,3} W(t) P(t)
$$
$$
+ b_{\alpha_2,5} W(t)^2 P(t) + b_{\alpha_2,6} P(t)^2
$$
$$
\hat{\alpha}_3(P, W, t) = b_{\alpha_3,0} + b_{\alpha_3,3} W(t) P(t) + b_{\alpha_3,7} P(t)^2 W(t)
$$
$$
\hat{\alpha}_4(P, W, t) = b_{\alpha_4,0} + b_{\alpha_4,3} W(t) P(t) \tag{19}
$$

Finally, the direct regression mean from Equation (15) is now completed as:

$$
\hat{\mu}_x(P, W, t) = b_{\mu,0} + b_{\mu,1} P(t) + b_{\mu,4} W(t)^2
$$
$$
+ b_{\mu,5} W(t)^2 P(t) \tag{20}
$$

Substituting the partial derivatives from Table 1 representing the coefficients $b_{0...7}$ into Equations (17), (18), (19), and (20) indicate that each statistic represents a partial differential equation. Moreover, the PDF $\hat{p}_x(x, P, W, t)$ itself is the solution to a partial differential equation representing the advective-dispersive transport of residential water demand in the three spatial dimensions x, P, W as well as time t.

Transport model application

Figure 4 depicts the late fall/winter November/December bimonthly histogram data from 2008, 2010, 2012, and
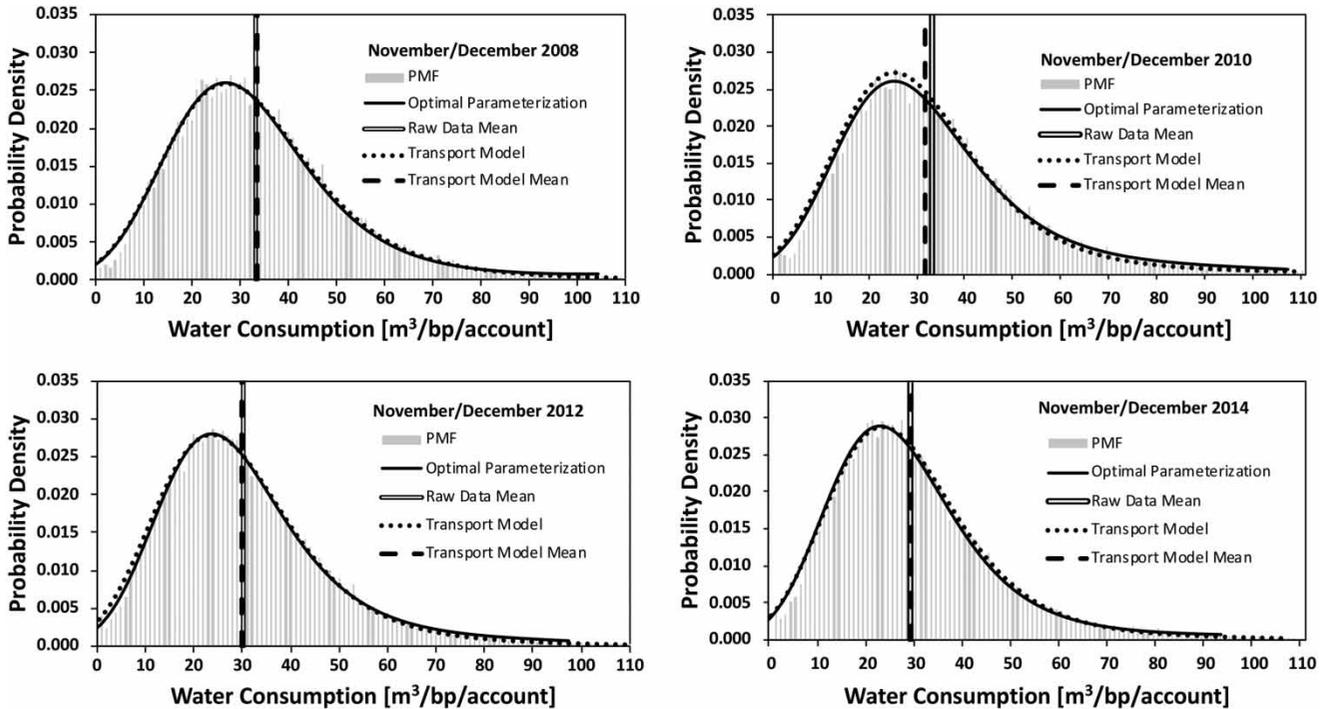
Figure 4 │ Transport model results for select November/December periods. Residential water consumption PMFs and corresponding PDFs for a sequence of November/December billing periods, with the optimal PDF from fitting the data, and PDF $\hat{p}_x$ obtained using the transport model. Also shown is the discrete mean $\mu_{x,i,\ell}$ as well as the estimated mean from the transport model $\hat{\mu}_{x|p_x}$.

2014 (see Figure 1) normalized into PMFs. Additionally, the parametric PDF arising from the optimal parametrization $p_{x,\ell}$ as well the advective-dispersive transport solution for $\hat{p}_x(x, P, W, t)$ given by Equation (17) are superimposed onto the PMFs. Finally, the transport mean $\hat{\mu}_{x|p_x}(P, W, t)$ given by Equation (18) as well as the arithmetic mean of the raw data $\mu_{x,i,\ell}$ are also presented. Figure 5 shows the same sequence of information except for the summer July/August bimonthly periods from 2007, 2009, 2011, and 2013. The advective-dispersive transport model for $\hat{p}_x(x, P, W, t)$ almost exactly reproduces the continuum response of residential water demand for the November/December periods which exhibit a low weather score, over the full range of real water price. The transport model becomes less accurate for the July/August summer months that correspond to a high weather score.

Figures 4 and 5 clearly indicate that as long as the continuum response of the entire system is adequately represented by $\hat{p}_x(x, P, W, t)$, then there is a unique representation for the transport mean $\hat{\mu}_{x|p_x}(P, W, t)$ such that it reproduces the arithmetic mean of the raw data. The

direct regression mean $\hat{\mu}_x(P, W, t)$ given by Equation (20) does not include any information regarding the shape of continuum response (using the control function parameters) and is derived by observing how the arithmetic mean of the raw data $\mu_{x,i,\ell}$ directly responds to P_ℓ and W_ℓ. This independence of the control function provides an additional avenue for verification of the advective-dispersive transport process by comparing $\hat{\mu}_{x|p_x}$ with $\hat{\mu}_x$, as well as against $\mu_{x,i,\ell}$ for all sampling periods ℓ. Values of $\hat{\mu}_{x|p_x}$ and $\hat{\mu}_x$ are visualized as a time series in Figure 6. Notice that the mean water consumption for July/August 2007 is somewhat underestimated by the transport model, perhaps due to the issue of averaging short-duration extreme summer weather events over a two-month period to quantify the ambient process of weather score W_ℓ. This is also observed during the 2010, 2012, and 2016 July/August bimonthly periods. However, both $\hat{\mu}_{x|p_x}$ and $\hat{\mu}_x$ exhibit nearly identical behavior for all sampling periods. The transport mean $\hat{\mu}_{x|p_x}$ is calculated by combining the location, scale, and shape of the continuum response as independent processes that all depend on $P(t)$ and $W(t)$. In contrast, the direct regression mean $\hat{\mu}_x$ does not differentiate
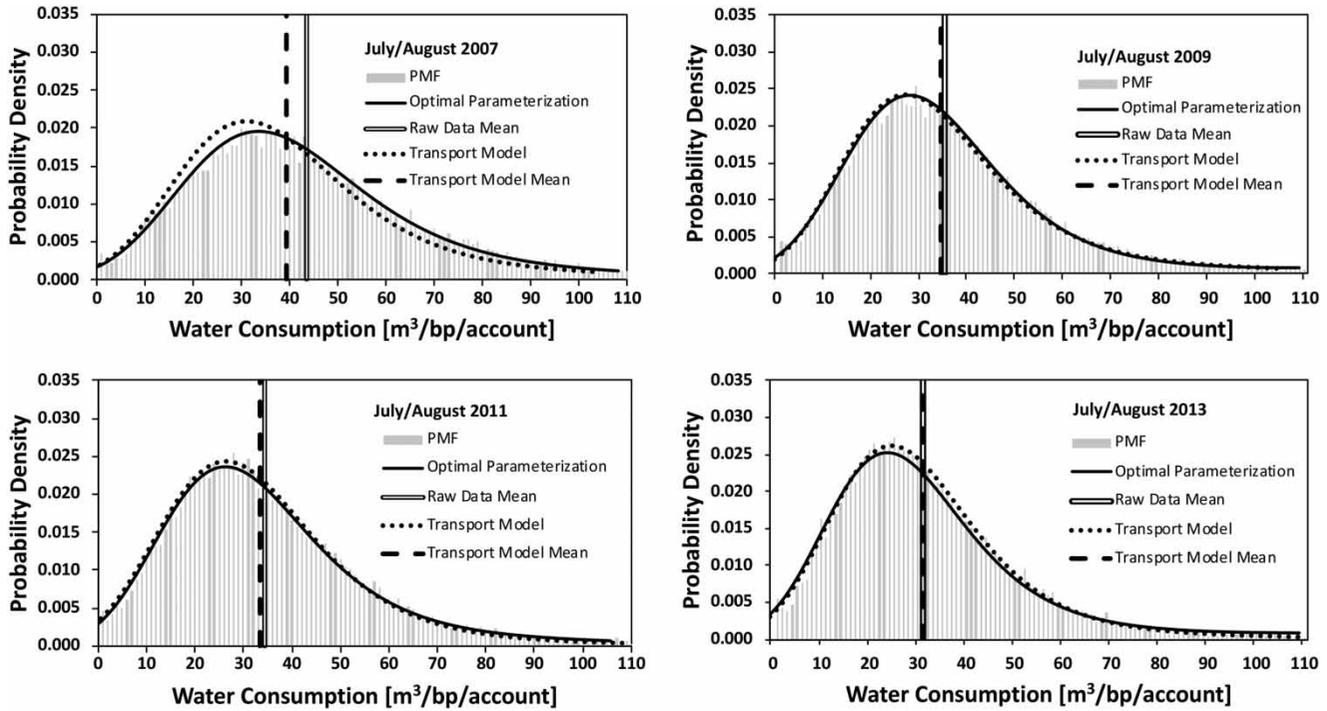
Figure 5 | Transport model results for select July/August periods. Residential water consumption PMFs and corresponding PDFs for a sequence of July/August billing periods, with the optimal PDF from fitting the data, and the PDF $\hat{p}_x$ obtained using the transport model. Also shown is the discrete mean $\mu_{x,j,\ell}$ as well as the estimated mean from the transport model $\hat{\mu}_{x|p_x}$.
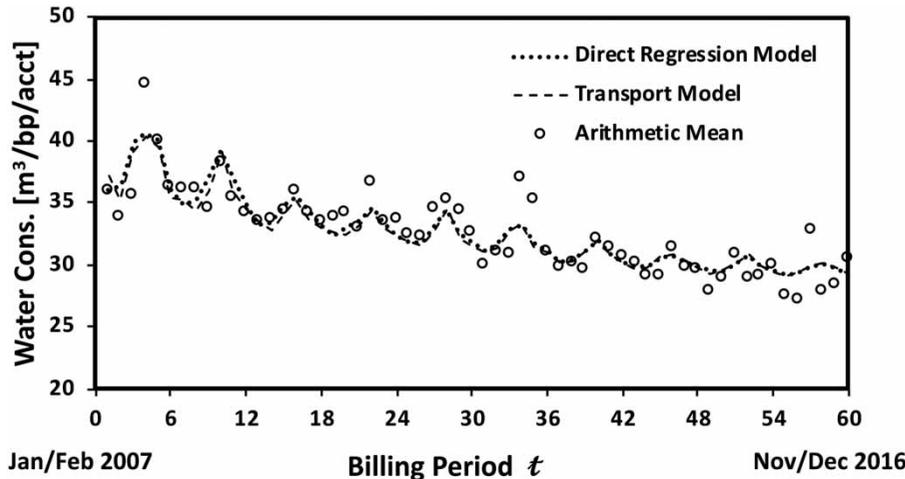

Figure 6 | Transport model results for mean water consumption. The measured mean statistics $\mu_{x,j,\ell}$, direct regression model results $\hat{\mu}_x$, and the corresponding transport model results $\hat{\mu}_{x|p_x}$ for the entire analysis period. See Appendix B, Tables B5 and B6 (available with the online version of this paper).

between the location, scale, and shape as it reproduces only the magnitude of the continuum response as a function of $P(t)$ and $W(t)$. Therefore, an analysis that only considers the direct regression model to assess a system's response naturally implies a loss of information.

PMFs from the summer billing periods of July/August 2015, May/June 2016, and July/August 2016 appear significantly different than previous years and indicate that omitted variables may be influencing water consumption. These are the only years where the measured mean water
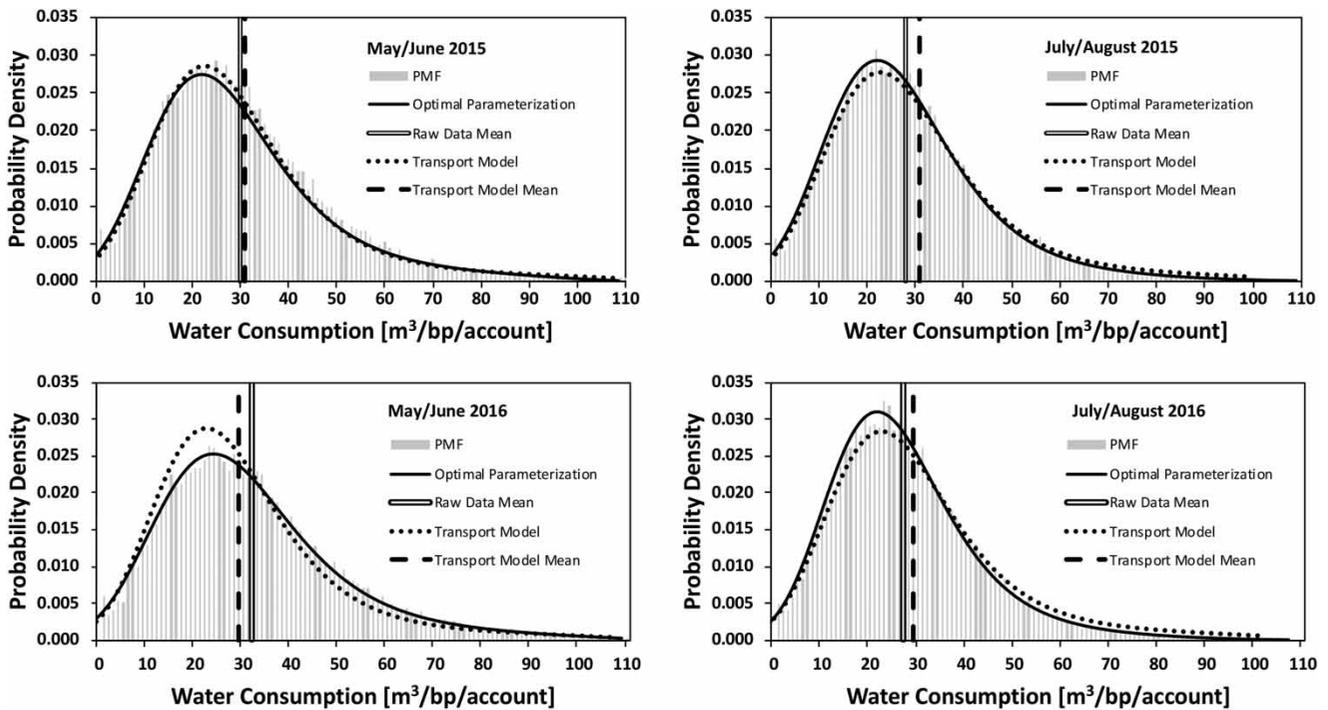
Figure 7 | Transport model results for 2015/2016 May/June and July/August periods. Residential water consumption PMFs and corresponding PDFs for a sequence of May/June and July/August billing periods, with the optimal PDF from fitting the data, and the PDF $\hat{p}_x$ obtained using the transport model. Also shown is the discrete mean $\mu_{x,i,\ell}$ as well as the estimated mean from the transport model $\hat{\mu}_{x|p_x}$.

consumption in May/June (30.86 $m^3/bp/acct$ for 2015 and 32.88 $m^3/bp/acct$ for 2016) is higher than the water consumption in July/August (28.96 $m^3/bp/acct$ for 2015 and 27.95 $m^3/bp/acct$ for 2016) of the same year. Figure 7 shows that the transport model $\hat{p}_x(x, P, W, t)$ overestimates the water consumption for the July/August period in both 2015 and 2016 despite the fact that the optimal parametrization $p_{x,\ell}$ is accurate. The hypothesis is that identifying and quantifying these potentially omitted variables could allow $\hat{p}_x(x, P, W, t) \rightarrow p_{x,\ell}$.

Starting in 2006, the Region of Waterloo implemented a *Water Efficiency Master Plan 2007–2015* which prioritized mass media advertising and a water conservation by-law that included outdoor water usage restrictions between May 31st and September 30th (Region of Waterloo 2006a, 2006b). Then, in 2015, the Region of Waterloo extended their water conservation practice with the *Water Efficiency Master Plan 2015–2025* including advertisement and enforcement. Additionally, they began to actively identify and target heavy water users through the Residential Water Savings Assistance Program (RWSAP):

'…*who are known to have especially high household water use will be actively contacted and encouraged to participate in the program. This program will help address the challenge shown frequently in market research that many residents are unaware that their consumption is markedly higher than the norm*' (Region of Waterloo 2014).

Beginning in May/June 2015, there appear to be tangible, yet unexplained, changes in residential water consumption habits that may indicate that water conservation education and by-law enforcement act as ambient processes that significantly confound the influence of price and weather. For instance, the mean water consumption during the May/June billing period in 2015 was the highest since 2011 and actually increased from 2015 to 2016. Also, both the median and standard deviation were higher in May/June 2016 and were lower in July/August of the same year relative to what the advective-dispersive transport model could predict. This indicates the model under-anticipated consumption in May/June and over-anticipated

consumption in July/August. Therefore, additional ambient processes arising from RWSAP could include water conservation arising from improved education, awareness through mass media advertising, and active enforcement of water restriction by-laws may have led individual residential accounts to decrease their water usage during the July/August billing period. However, residents also appear to have substituted their water usage by increasing their water consumption in advance of the May 31st water restriction deadline prior to decreasing their consumption in the July/August period. Thereafter, even though they are conforming to policy, residential consumers return to historical water consumption patterns for the fall and winter months.

In a specific demonstration of the continuum analogy for water consumption, this framework is able to track the probability that a residential account exists below a water poverty threshold or uses excessive water within the analysis period. According to the 2016 census bulletin (Region of Waterloo 2016), the average household has three occupants. The World Health Organization recommends a minimum of 20 liters of water per occupant per day for survival (WHO 2013). Using a threshold of 7 $[m^3/bp/acct]$ (approximately double the survival minimum), the advective-dispersive transport model may provide a resource to track and anticipate the percentage of households at risk for water poverty. In 2007, approximately 2.7–3.0% of households used less than 7 $[m^3/bp/acct]$ which then subsequently increased to approximately 4.5–6.0% between 2014 and 2016. As the water consumption PDF compresses to the origin, the number of households limiting their water consumption increases. It is interesting to note that the percentage of households below 7 $[m^3/bp/acct]$ was consistent between 2014 and 2016 after the water use by-law came into effect. It is unlikely that these households were contacted and may not have been influenced by the RWSAP. The advective-dispersive transport model can also track and anticipate the threshold of the top 10% of water users to see if price, policy, or weather has an impact. The upper 10% consumed greater than approximately 60 $[m^3/bp/acct]$ in the winter and 77 $[m^3/bp/acct]$ in the summer of 2007. By 2014, their consumption decreased to 50 $[m^3/bp/acct]$ in the winter and 55 $[m^3/bp/acct]$ in the summer. After the RWSAP was established in 2016, their consumption became almost uniform at 47 $[m^3/bp/acct]$ in the winter and summer; however, these

consumers used approximately 53 $[m^3/bp/acct]$ for the November/December period indicating there may be some holiday season effects. Clearly, there is some relationship between price, weather, policy, and water consumption. It is difficult to include all potential influences or even isolate those that intuitively should affect household water consumption. The advective-dispersive transport model may provide an additional tool to quantify and understand these complex interactions and allow water utilities to better plan for the future. With the introduction of additional sophistication to quantify the impact of policy initiatives, future iterations of the advective-dispersive transport model could provide a foundation for defensible and accurate forecasting.

CONCLUSIONS

The motivation for this study was to develop a methodology for water utilities to understand how changes to water price, weather patterns, and municipal policy initiatives influence the relative probability that households will be water impoverished or excessively consume water. To replicate the observed response of the water consumption histogram, this analysis individually fits a parametric PDF for each bimonthly period using a third-order exponential polynomial control function and reproduces the water consumption distribution as a continuous function through time. The resulting PDFs represent the solution to an advection-dispersion transport equation, where: the median represents advection by locating the PDF and the standard deviation combined with the standard-score space PDF represents dispersion by virtue of giving the solution scale and shape. Optimal parameterization of the transport model requires conservation of probability, which replicates shifts in the distribution as a one-dimensional advective-dispersive transport process along the axis of water consumption x $[m^3/bp/acct]$. Therefore, the probability $\mathcal{P}$ of residential water consumption occurring within any prescribed interval $\Delta x = x_{max} - x_{min}$ where $\mathcal{P} = \int_{x_{min}}^{x_{max}} p_x(x, P, W, t)dx$ under predefined ambient conditions of real water price $P(t)$ and weather score $W(t)$ at time t.

The outcome of this analysis provides new possibilities for interpreting how the location, scale, and shape of a

distribution of measurements respond to changes in ambient conditions. This analysis demonstrates that it is reasonable to disaggregate the data into advection and dispersion components to characterize how a distribution will evolve through time. This approach could provide utilities with: the ability to quantify the influence of policy implementation and enforcement such as summer water use restrictions; and also implement policy to target populations including water impoverished households and excessive consumers within the utility. Although this approach requires a certain degree of competency in mathematics and programming to implement, the virtue of this methodology is that it is conducive to automation. Additionally, this advective-dispersive transport framework for consumer behavior could analogously apply to other industries such as electrical utilities and transportation services, as well as social and health science applications with histogram data that exhibit non-Gaussian tendencies.

REFERENCES

Donkor, E. A., Mazzuchi, T. A., Soyer, R. & Roberson, J. A. 2011 Urban water demand forecasting: review of methods and models. *Journal of Water Resources Planning and Management* **140** (2), 146–159.

Enouy, R. 2018 *An Investigation Into Water Consumption Data Using Parametric Probability Density Functions*. UWSpace, PhD thesis, University of Waterloo, Canada. http://hdl.handle.net.proxy.lib.uwaterloo.ca/10012/13545.

Environment Canada 2017 *Kitchener/Waterloo Weather Station Historical Data*. http://climate.weather.gc.ca/historical_data/search_historic_data_e.html (accessed June 2016).

Environmental Protection Agency (EPA) 2003 *Water and Wastewater Pricing – An Informational Overview*. Office of Wastewater Management. EPA 832-F-03-027.

EPA 2005 *Case Studies of Sustainable Water and Wastewater Pricing*. Office of Water. EPA 816-R-05-007.

EPA 2006 *Expert Workshop on Full Cost Pricing of Water and Wastewater Service*. November 1–3, 2006. Michigan State University, Institute for Public Utilities. Office of Water. EPA 816-R-07-005.

Gargano, R., Tricarico, C., Granata, F., Santopietro, S. & de Marinis, G. 2017 Probabilistic models for peak residential water demand. *Water* **9**, 417. doi:10.3390/w9060417.

Gauss, C. F. 1816 Bestimmung der Genauigkeit der Beobachtungen (Determination of the accuracy of observations).

Zeitschrift für Astronomie und Verwandte Wissenschaften **1**, 187–197.

Hogg, R. V. & Craig, A. T. 1995 *Introduction to Mathematical Statistics*, 5th edn. Macmillan, New York, USA.

House-Peters, L. A. & Chang, H. 2011 Urban water demand modeling: review of concepts, methods, and organizing principles. *Water Resources Research* **47**, 15.

Hughes, J. A. & Leurig, S. 2013 *Assessing Water System Revenue Risk: Considerations for Market Analysts*. Environmental Finance Center and Ceres, University of North Carolina, Chapel Hill, NC, USA.

McGranahan, G. 2002 *Demand-side Water Strategies and the Urban Poor*. International Institute for Environment and Development (IIED), PIE Series No. 4.

Ministry of the Environment (MOE) 2011 *Water Opportunities and Water Conservation Act, 2010*. Ministry of the Environment, Ontario. Environmental Registry Number 010e9940, http://www.e-laws.gov.on.ca/html/source/statutes/english/2010/elaws_src_s10019_e.htm.

National Aeronautics and Space Administration (NASA) 2016 *NASA Langley Research Center POWER Project*. Latitude: 43.464, Longitude: 80.52. https://power.larc.nasa.gov/cgi-bin/agro.cgi (accessed June 2016).

Pearson, K. 1894 On the dissection of asymmetrical frequency curves. *Philosophical Transactions of the Royal Society A* **185**, 71–110. doi:10.1098/rsta.1894.0003.

Region of Waterloo 2006a *Water Efficiency Master Plan Update 2007–2015*. Transportation and Environmental Services – Water Services. Available online: http://www.regionofwaterloo.ca/en/aboutTheEnvironment/resources/water%20efficiency%20master%20plan.pdf.

Region of Waterloo 2006b *Environews*. April 2006. p. 8. Available online: http://www.regionofwaterloo.ca/en/aboutTheEnvironment/resources/Spring06.pdf (accessed July 2017).

Region of Waterloo 2014 *Water Efficiency Master Plan 2015–2025*. Department of Water Efficiency. Available online: http://www.regionofwaterloo.ca/en/aboutTheEnvironment/resources/WEMP_2015_-_2025_FINAL__May_1_2025.pdf.

Region of Waterloo 2016 *Census Bulleting: Households, Families, and Marital Status*. Region of Waterloo.

University of Waterloo 2016 *University of Waterloo Weather Station Historical Data*. http://weather.uwaterloo.ca/data.html (accessed June 2016).

World Health Organization (WHO) 2013 How Much Water is Needed in Emergencies? Technical Notes on Drinking-water, Sanitation, and Hygiene in Emergencies. In: *Prepared by Water, Engineering and Development Centre (WEDC) for WHO*.

Zambom, A. Z. & Dias, R. 2012 A review of kernel density estimation with applications to econometrics. *International Econometric Review* **5**, 20–42.