# Water Demand Forecasting Model Application

by

Junhao Lu

A thesis

presented to the University Of Waterloo in

fulfilment of the

thesis requirement for the degree of

Master of Earth Science in

Earth Science

Waterloo, Ontario, Canada, 2019

# AUTHOR DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

# ABASTRACT

Forecasting water demand requires quantifying potential relationships between relevant statistics and ambient conditions such as water price and weather. Dr. Enouy (2018) demonstrates that discrete histograms can be parameterized into continuous probability density functions. Consistent parametrization allows regression analysis to be applied to the PDF statistics, thus able to reproduce PDFs through time.

This work briefly introduces Dr. Enouy's (2018) methodology and mainly investigates the applicability of this method. It formalizes the implementation details of residential water application in terms of data culling, optimization and regression analysis. A modified version of this method is employed as an adaptation to the analysis of commercial water demand.

This thesis also discusses the possibility of employing the scheme of software development, to assure the robustness and correctness of this implementation.

# ACKNOWLEDGEMENTS

I would like to express my very great appreciation to Dr. Andre Unger, my research supervisors, for his patient guidance and useful critiques of this research work, and Dr. Robert Enouy for their valuable suggestions during the planning and development of this research work.

# Table of Contents

# LIST OF TABLES

# LIST OF FIGURES

# 1 INTRODUCTION

Our water distribution systems were constructed without consideration of the necessity for future maintenance. For instance, most of the watermain and sanitary sewer infrastructure in Waterloo Region are between 30-50 years old with some even more than 60 years old. According to the AWWA report Buried No Longer, most of these water infrastructures developed in the post-WWII are approaching the end of their service life. To maintain current levels of water service, at least $1 trillion is estimated to be spent on maintenance and replacement of these systems over the next 2 decades in the US (AWWA 2012). Faced with increased costs of operations and maintenance, higher revenues need to be generated by raising water prices.

Tap water sales has been in decline for more than a decade, which leads to a large downfall on revenues for the City of Waterloo. In order to generate enough revenues to cover expenses to maintain the water distribution system, cities are forced to raise their water rates. However, failure to correctly quantify the decline in consumer demand resulting from price increases can even lead to an even greater shortfall in revenues. Therefore, to develop an accurate financial model is important for anticipating how consumer demand responds to price changes and other factors such as weather.

The new financial model generates continuous probability density functions (PDFs) from the water demand histogram from historical data, following the methodology developed by Enouy (2018). These data were obtained from the City of Waterloo for years 2007 through 2014, and represented bimonthly water consumption of both residential and commercial accounts. The objective of this work is to validate and formalize the methodology developed by Enouy (2018) into a scalable object-oriented software algorithm and SQL database capable of quantify the demand response of any large city. This objective is realized by analyzing both the residential and commercial water demand response of the City of Waterloo.

Variability in the location, scale and shapes of the water consumption PDFs over time indicate a relationship between changes in price and weather, which is represented by temperature and rainfall, and shifts in the consumer demand. This work uses the computational methodology to quantifying the relationship between the control function parameters that define the PDF, as well as the median, standard deviation and mean statistics to that of water price, weather, household income and demographics. The outcome is to be able to forecast residential, commercial, institutional, and industrial water demand for the City of Waterloo under anticipated water price and weather scenarios.

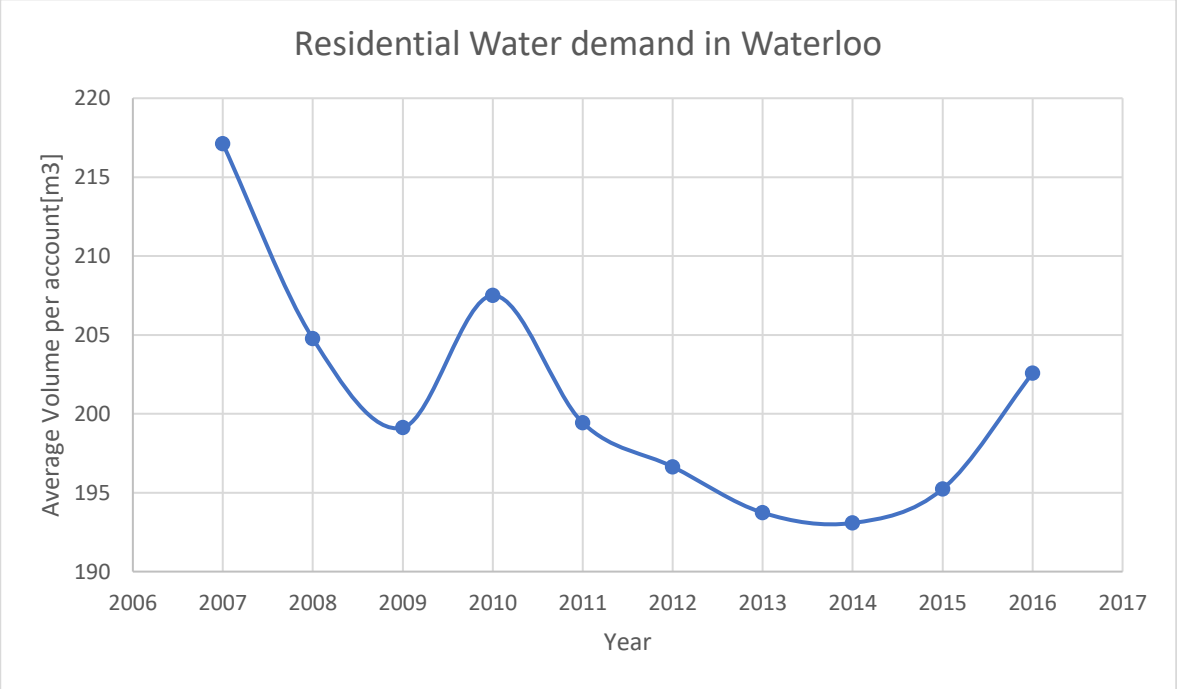



**Figure 1 Average residential water consumption for last 10 years**

The following three steps broadly outline the scope of the software development and implementation. First, apply the methodology from Enouy (2018) to transforming water consumption histogram into continuous probability density functions, with a tabulated list of control function parameters as a function of ambient price and weather score. Second, perform multi-variate curvilinear regression on the median, standard deviation, and control

function parameters as a function of the dependent variables of water price and weather score (as defined by precipitation and temperature). Third, reproduce probability density functions that can be used to calculate the probability of achieving a threshold water demand as a function of price and weather.

# 2 THEORY

The foundation of this work is based on Enouy (2018), and in particular his method of reproducing discrete histogram data as a continuously differentiable parametric PDF, and quantify how the continuum statistics of a PDF evolve as a function of multiple ambient processes. This section briefly introduces the most important aspects of his methodology, as a guideline of the subsequent section for applications.

## 2.1 Median and Standard Deviation

The importance of median and standard deviation statistics needs to be addressed here for the discussion of statistical transformation in the subsequent section. The central tendency of the histogram of any discrete dataset can be represented by its median $m_{x,i}$ , while the scale is represented by its standard deviation $\sigma_{x,i}$ (Enouy, 2018). The definition of standard deviation is modified for the purpose of this particular analysis is:

$$\sigma_{x,i} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}[x_i - m_{x,i}]^2} \tag{2.1}$$

Where, $N$ is the total number of measurements, and $x_i$ is the ith measurement.

## 2.2 Continuously Differentiable PDFs

This section discusses transformation from a PMF $p_{x,k}$, a functional representation of a histogram, to a continuously differentiable PDF $p_x$. First, $p_{x,k}$ needs to be expressed in the standard-score space as $p_{z,k} = \sigma_{x,i} p_{x,k}$. This transformation will be specifically defined in the following section. The corresponding $p_z$ is characterized by a control function $g_z$, which represents the lognormal derivative of $p_z$.

$$g_z = \frac{1}{p_z}\frac{dp_z}{dz} \tag{2.2}$$

$$\Rightarrow p_z = \exp(\int g_z \, dz) \tag{2.3}$$

By optimizing the control function to fit $p_{z,k}$, we are able to match the shape of the PMF. Meanwhile, guarantee the integral of $p_z$ on our definite scale to unit area (Enouy, 2018).

$$c_z = \int_{z_0}^{z_1} p_z dz \tag{2.4}$$

Now combined with the location, scale represented by median and standard deviation, the three attributes of a PDF are captured.

## 2.3 Statistical Transformations

Enouy (2018) introduce a new transformation referred to as median-relative space as a way to normalize PMFs and PDFs by dividing each measurement by the median statistic. A key attribute to making this projection possible, is to make sure that the CDF and the integration of PDF are identical:

$$\int p_x^* \, dx = \int p_y^* \, dy = \int p_z \, dz \tag{2.5}$$

Where, $p_x^*$, $p_y^*$, and $p_z$ represent the zero-centered PDFs in the measurement, median-relative, and standard-score spaces. The * superscript represents a distribution centered at zero by subtracting a median value in the measurement space $x^* = x - m_x$ (Enouy et al., 2018). Table 2.1 introduces the transformations for continuous zero-centered PDFs between each spatial representation.

**Table 2. 1 Spatial transformation in measurement, median-relative and standard-score space**

| Space | Magnitude | PDF | Derivative |
|---|---|---|---|
| $\boldsymbol{x}$ | $x_i$ | $p_x^* = \frac{1}{m_{x,i}} p_y^* = \frac{1}{\sigma_{x,i}} p_z$ | $dx = m_{x,i} dy = \sigma_{x,i} dz$ |
| $\mathbf{y}$ | $y_i = \frac{x_i}{m_{x,i}} = \frac{\sigma_{x,i}}{m_{x,i}} z_i + 1$ | $p_y^* = \frac{1}{\sigma_{x,i}} p_y^* = m_{x,i} p_z$ | $dy = \frac{\sigma_{x,i}}{m_{x,i}} dz = \frac{1}{m_{x,i}} dx$ |
| $\mathbf{z}$ | $z_i = \frac{y_i - 1}{\frac{\sigma_{x,i}}{m_{x,i}}}$ | $p_z$ | $dz = \frac{m_{x,i}}{\sigma_{x,i}} dy = \frac{1}{\sigma_{x,i}} dx$ |

## 2.4 The Control Function

As mentioned in the previous section, control function is defined as lognormal derivative of a PDF in standard-score space. Equation 2.6 is the basic form of a first-order control function that produces a normal distribution.

$$g_z = -[\alpha_1 + \alpha_2 z] \tag{2.6}$$

Where $\alpha_1$ and $\alpha_2$ represent the control function parameters.

Equation 2.7 represents a polynomial series expansion for the control function. It is reshaped by adding additional polynomial terms to a normal distribution (Enouy, 2018) as shown in the following equations.

$$g_z = -\left[\alpha_1 + \tan\left(\frac{\alpha_2 \pi}{180}\right) z + \sum_{n_z=1}^{N_z} \alpha_{n_z+1} z^{n_z+1}\right] \tag{2.7}$$

where $\alpha_{n_z}$ is the parametric constant.

$n_z$ represents the order on the standard-score variable $z$.

$N_z$ is the total order of the control function in the standard-score space.

Equation 2.8 presents the measurement space PDF $p_x$ as a projection of the standard-score PDF $p_z$ using a combination of the median $m_x$ and standard deviation $\sigma_x$ (Enouy, 2018). It is applied in the notion of statistical advection and dispersion:

$$p_z = f(g_z), \qquad g_z = f\left(a_{n_z}\right) \tag{2.8}$$

$$\underbrace{p_x}_{\substack{Continuum \\ Distribution}} = \underbrace{m_x}_{\substack{Advective \\ Process}} + \underbrace{\frac{1}{\sigma_x} \times p_z}_{\substack{Dispersive \\ Process}} \tag{2.9}$$

where,

$g_z$ is a polynomial series control function that represents the lognormal derivative of the standard-score PDF;

$a_{n_z}$ represent the shape parameters that characterize the control function; and,

$n_z$ is the number of terms in the polynomial series.

## 2.5 Median-Relative Space

The introduction of the median-relative space is a very important feature of Enouy (2018). It provides an effective means of normalizing datasets for the purposes of data culling, which significantly improves parameter estimation when fitting the PDF for different sets and scales of histogram data. Its value in the context of analyzing the City of Waterloo residential and commercial data is described in the following sections.

### 2.5.1 Data Culling

For many discrete datasets, even after being transformed into the median-relative space, the upper bound could still be infinitely large. That small portion of data values at the tail could significantly disrupt the parameter estimation process for the control function defining the system. Therefore, these outliers were discarded from the dataset (Enouy, 2018). Excluding those data did not have a great influence on the median statistic. However, it did significantly affect the fitting processes for estimated control function parameters as well as the standard deviation statistic. Therefore, a proper upper bound ought to be set in the median-relative as $y_{max}$, culling ratio. The setting of culling ratio will be further discussed in Section 3.1.

### 2.5.2 Objective Function

The objective function that needs to be minimized as a least-square problem is defined in the following form:

$$MSE_{c,y} = \frac{1}{N_k}\sum_{k=1}^{N_k}[c_y - c_{y,k}]^2 \tag{2.10}$$

Where $N_k$ represents the number of bins

$c_y$ represents CDF probability at $y$

$c_{y,k}$ represent CMF probability at $k$th bin

the median-relative space allows the size of each probability interval bin be independently predefined for all different datasets.

### 2.5.3 The Mean Statistic

The probability weighted mean $\mu_x$ is defined as:

$$\mu_z = \int_{z_{min}}^{z_{max}} z p_z \, dz \tag{2.11}$$

$$\mu_x = m_{x,i}\mu_y = m_{x,i} + \alpha_{x,i}\mu_z \tag{2.12}$$

The arithmetic mean $\mu_{y,i}$ is compared to $\mu_y$ as another standard to evaluate the quality of a fitting.

$$MSE_{\mu,y} = \left[\mu_y - \mu_{y,i}\right]^2$$

## 2.6 Curvilinear Regression

The goal of this model is to be able to quantify how both the residential and commercial water demand is impacted by multiple ambient processes, such as price, temperature, precipitation, water restriction by-law enforcement, water conservation policies (Enouy, 2018). This could be achieved by predicting how the optimal fitted PDF shifts as a response to the processes listed above. However, processes like restrictions and policies are unlikely to be quantified. We can only focus on the tangible processes of price, temperature, precipitation. Furthermore, we assume that the intangibles are minimal compared to others, and their impact can be inferred via changes in the water consumption that could not be explained by the other three major factors (Enouy, 2018). Prices are adjusted using the annual consumer price index (CPI) inflation rate to a base year of 2004. Weather is represented as a combined score of temperature and precipitation as:

$$W = T \times R$$

where $T$ represents the average of the daily high temperature in degrees Celsius for all days within sampling periods (University of Waterloo Weather Station, 2017) . Precipitation $R$ represents the number of days with less than 2mm of rainfall during sampling periods (NASA, 2017; Environment Canada, 2017). The weather score is suggested by Enouy (2018) to avoid the impact brought by the inter-dependence between temperature and precipitation.

Let $P$ and $W$ represent real price and weather score. Let $U$ represent $\{\mu_x, m_x, \sigma_x \alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ which are all the parameters that we regress with price and weather. The total differential of $dU$ can be expressed as:

$$dU = F_{u,p}dp + F_{u,w}dw \tag{2.13}$$

Where $F_{u,p}(p,w) = \frac{\partial \mathrm{U}}{\partial \mathrm{p}}$

$F_{u,w}(p,w) = \frac{\partial \mathrm{U}}{\partial \mathrm{w}}$

$F_{U,p}$ and $F_{U,w}$ can be expanded using a Taylor series expansion around p = 0 and w = 0:

$$\begin{aligned} F_{U,p} - F_{U,p=0} = {} & \frac{\partial F_{U,p}}{\partial w}dw + \frac{\partial F_{U,p}}{\partial p}dp + 2\frac{\partial^2 F_{U,p}}{\partial p \partial w}dpdw + \frac{1}{2}\frac{\partial^2 F_{U,p}}{\partial p^2}dp^2 \\ & + \frac{1}{2}\frac{\partial^3 F_{U,p}}{\partial p^2 \partial w}dp^2 dw + \cdots \\ F_{U,w} - F_{U,w=0} = {} & \frac{\partial F_{U,w}}{\partial p}dp + \frac{\partial F_{U,w}}{\partial w}dw + 2\frac{\partial^2 F_{U,w}}{\partial p \partial w}dpdw + \frac{1}{2}\frac{\partial^2 F_{U,w}}{\partial w^2}dw^2 \\ & + \frac{1}{2}\frac{\partial^3 F_{U,w}}{\partial w^2 \partial p}dw^2 dp + \cdots \end{aligned} \tag{2.14}$$

Next, substituting Equation 2.14 into 2.13 and compressing the notation using ${F_{U,p}}' = \frac{\partial F_{U,p}}{\partial p}$, ${F_{U,p}}'' = \frac{\partial^2 F_{U,p}}{\partial p^2}$, ${F_{U,p}}^* = \frac{\partial F_{U,w}}{\partial w}$, ${F_{U,p}}^{**} = \frac{\partial^2 F_{U,w}}{\partial w^2}$, ${F_{U,p}}^{*\prime} = \frac{\partial F_{U,p}}{\partial w}$, ${F_{U,p}}^{*\prime\prime} = \frac{\partial^2 F_{U,p}}{\partial p \partial w}$, ${F_{U,p}}'^* = \frac{\partial F_{U,w}}{\partial p}$, ${F_{U,p}}^{**\prime} = \frac{\partial^2 F_{U,w}}{\partial p \partial w}$, ${F_{U,p}}^{*\prime\prime\prime} = \frac{\partial^3 F_{U,p}}{\partial p^2 \partial w}$, ${F_{U,p}}^{***\prime} = \frac{\partial^3 F_{U,w}}{\partial w^2 \partial p}$ results in:

$$dU = [F_{U,p=0} + {F_{U,p}}^{*\prime}dw + {F_{U,p}}'dp + 2{F_{U,p}}^{*\prime\prime}dpdw + \frac{1}{2}{F_{U,p}}''dp^2 \tag{2.15}$$

$$+\frac{1}{2}F_{U,p}{}^{*'''}dp^2dw+\cdots]dp+[\,F_{U,w=0}+F_{U,w}{}^{'*}dp+\,F_{U,w}{}^{*}dw+2F_{U,w}{}^{**'}dpdw+\frac{1}{2}F_{U,w}{}^{**}dw^2+\frac{1}{2}F_{U,w}{}^{***'}dw^2dp+\cdots]dw$$

Enouy (2018) truncate higher-order terms of the Taylor series expansion to avoid overfitting, $F_{U,p}{}^{***'} = F_{U,p}{}^{*'''} = F_{U,p}{}^{**} = F_{U,p}{}'' = 0$, resulting in:

$$dU = [F_{U,p=0}+F_{U,p}{}^{*'}dw+\,F_{U,p}{}'dp+2F_{U,p}{}^{*''}dpdw+\,]dp+[F_{U,w=0}+\,F_{U,w}{}^{'*}dp+\,F_{U,w}{}^{*}dw+2F_{U,w}{}^{**'}dp]dw \quad (2.15)$$

For conditions where $dp = p - 0$ and $dw = w - 0$, integrate on U:

$$\int U = F_{U,p=0}p+\,F_{U,w=0}w+\left(F_{U,p}{}^{*'}+\,F_{U,w}{}^{'*}\right)pw+\frac{1}{2}F_{U,w}{}^{*}w^2+F_{U,w}{}^{**'}w^2p+F_{U,p}{}'p^2+F_{U,p}{}^{*''}p^2w \quad (2.16)$$

Finally, all partial derivatives from the Taylor series expansion can be treated as coefficients of a curvilinear regression model as:

$$U = b_0+b_1p+\,b_2w+b_3pw+b_4w^2+b_5w^2p+b_6p^2+b_7p^2w \quad (2.17)$$

# 3 APPLICATION

This section discussed the application of the above methodology for the analysis of residential and commercial water demand for the City of Waterloo. Water consumption data is obtained from 2007 to 2016. This section discuses pertinent details of implementation for some important processes, and compares the performance and quality of results obtained using a variety of numerical methods in the implementation of the algorithm.

## 3.1 Residential Water Demand Analysis

The City of Waterloo has a total 38,555 residential accounts created from 2007 to 2016, with each single billing period having more than 20,000 active accounts. The quantity of active accounts guarantee that the sequence of PMFs each exhibit a smooth and continuous shape representing a continuum response to transit price and weather conditions.

### 3.1.1 Data Culling

The first step in the data culling process is to examine the dataset from each billing period to determine its maximum value, with peak residential water demand values shown on Figure 3.1.

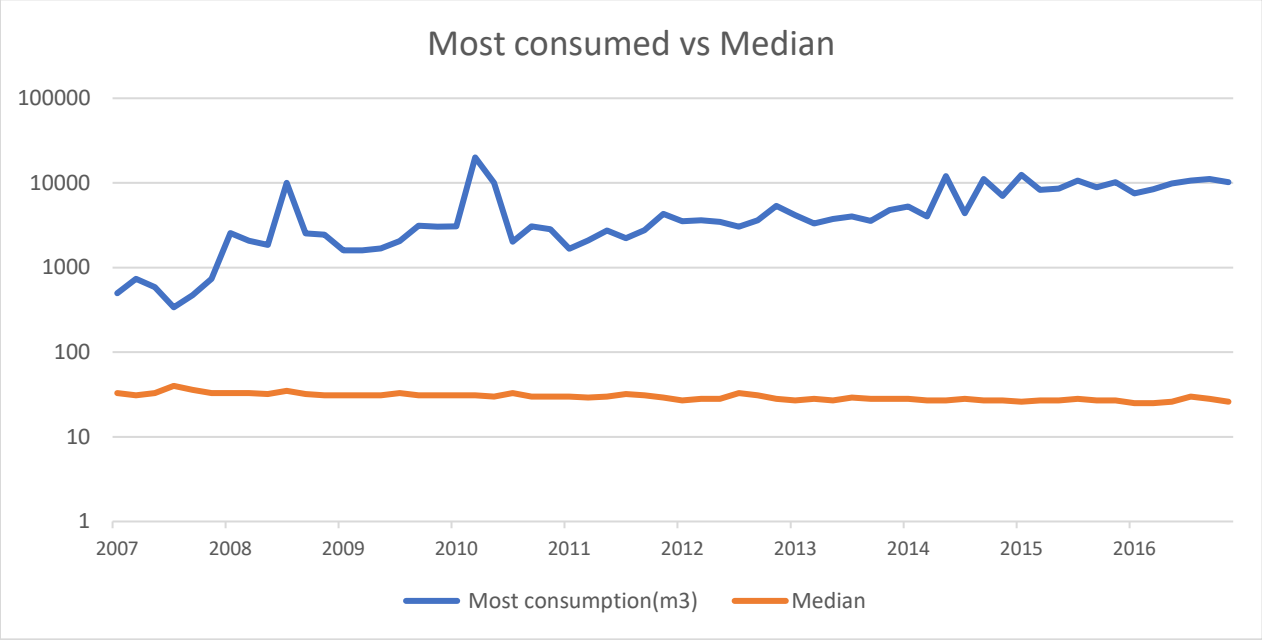



**Figure 3. 1 Max volume consumed by a single account with median of each billing period**

As can be seen from Figure 3.1, the maximum volumes consumed by as single account can be hundreds and thousands of times of its median in each billing period. This could result from excessive measurement error or perhaps observations from another distinct population (Enouy, 2017). For example, a student housing condo with thousands of individuals may be labeled as a single residential account by the City of Waterloo, but would obviously not represent the water consumption behavior of single family residence. These population outliers can potentially bias our evaluation of the median and standard deviation, as well as the parameters within the control function given their reliance on the standard-score space.

To remove those outliers, an upper bound in the median-relative space is predefined. The upper bound in the median-relative space was defined a-priori as $y_{max} = 4$ (Enouy, 2018). A key contribution from this code design is to specify an algorithm to determine a suitable value based on the shape of the histogram, subject to restricts defined by user input .

The implementation does not allow more than 5% of the data to be removed as part of the culling process. The algorithm begins with $y_{max} = 4$ as the upper bound. Two conditions are then checked:

1) that no more than 5% percent of the dataset has been culled; and,
2) the median did not shift by more than 10% percent relative to the old median.

If both conditions are met, then $y_{max} = 4$ is applied as a culling ratio. If the percentage of data removed decreases by more than 1%, then increase $y_{max}$ by increments of 1 until a final value of 10 is reached. Otherwise, four becomes the default culling ratio. If either of the two conditions are not met, then increase the ratio by increments of 1 and repeat the process above until 10 is reached. If the culling ratio reaches 10, the distribution is classified as "heavy-tailed" and is transformed into $\ln y$ space. This transformation is discussed in detail in the next section.

Pseudocode:

```
culling ratio <- 4
while (culling ratio < 10)
        data culling
        if (total after being culled /total < 95% and new median/old median > 90%))
                old culling rate <- 1 - total after being culled /total
                new culling rate <- 0
                while (|new culling rate - new culling rate| > 1% and culling ratio < 10)
                        culling ratio <- culling ratio + 1
                        data culling
                        new culling rate <- 1 - total after being culled/total
                break
        else
                culling ratio <- culling ratio
if culling ratio >= 10
   This is a heavy-tailed distribution
```

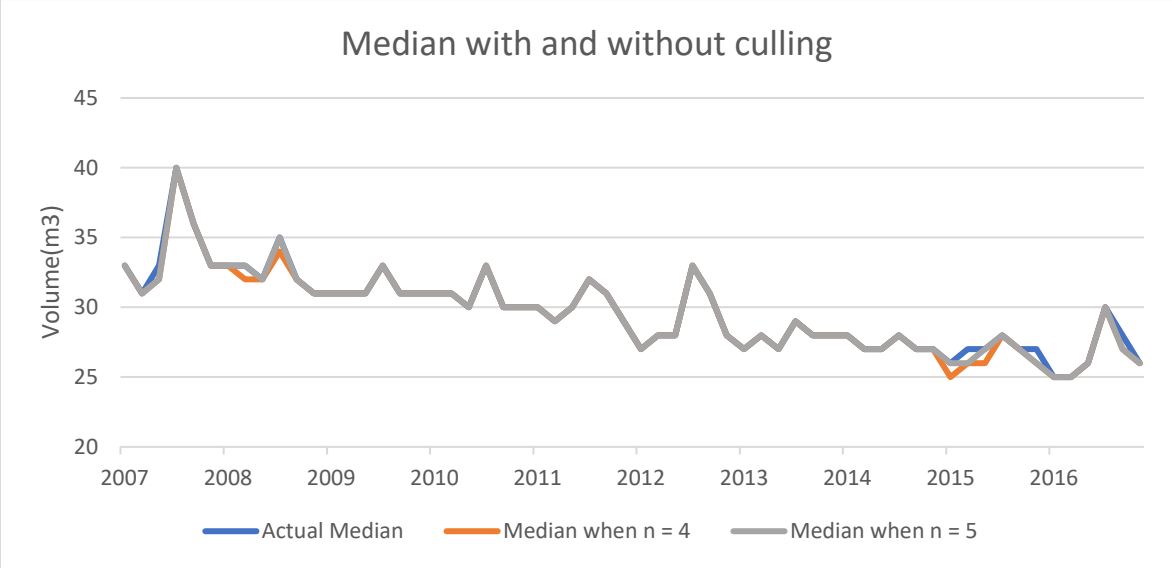



**Figure 3. 2 Median and median after culling by different ratio**

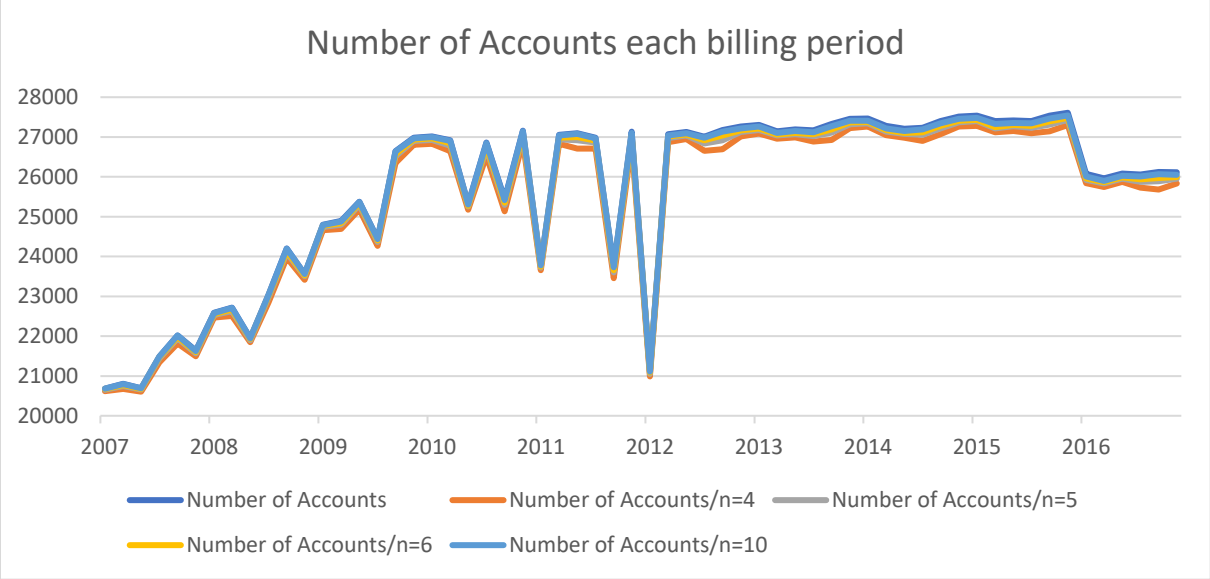



**Figure 3. 3 Number of accounts before and after being culled by different ratios**

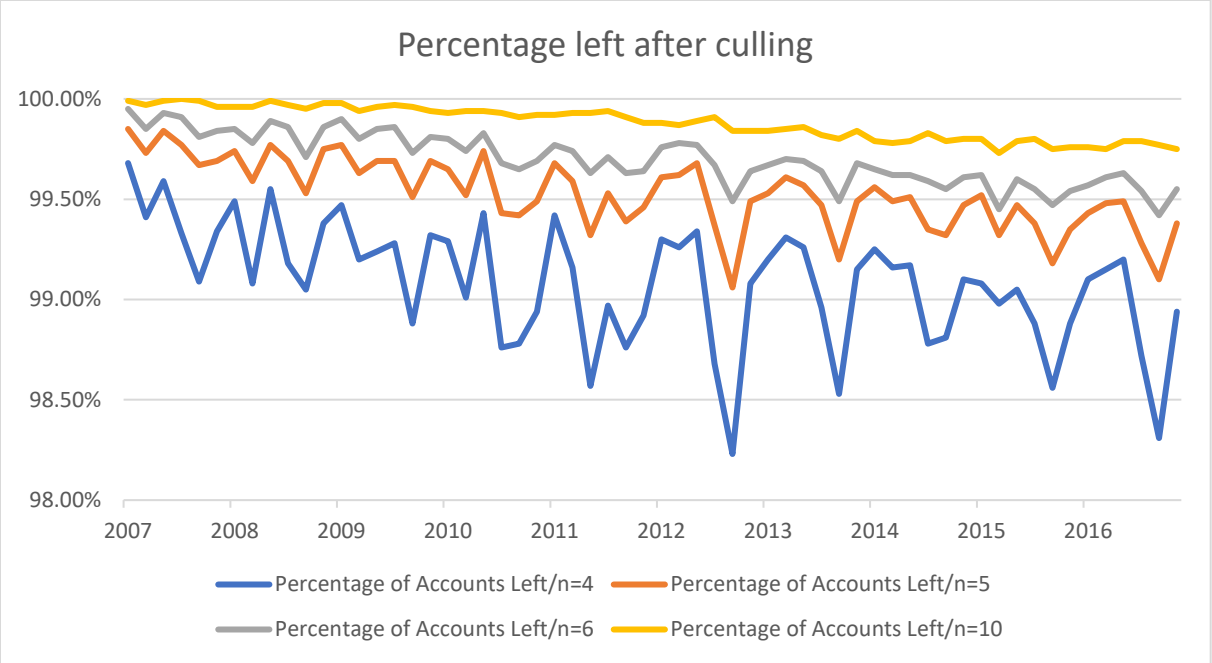



**Figure 3. 4 Percentage left after culling by different ratios**

When the culling factor is four, the median of most of the billing periods did not shift at all with some shifting at most by $1\ m^3/bp/account$. This guarantees a stable environment for removing population outliers without recursively shifting the median statistic (Enouy, 2018). Figure 3.3 and 3.4 demonstrates that in most of the billing periods, having 4 as the upper bound remove less than 2% percent of the dataset. This outcome of this culling method dramatically improved the convergence rate, accuracy, and the stability of the algorithm. Table 3.1 demonstrates that data culling reduced the number of non-linear Iterations to achieve convergence from 28 to 8 on average, while the residual error was reduced from $5.84 \times 10^{-4}$ to $2.98 \times 10^{-5}$.

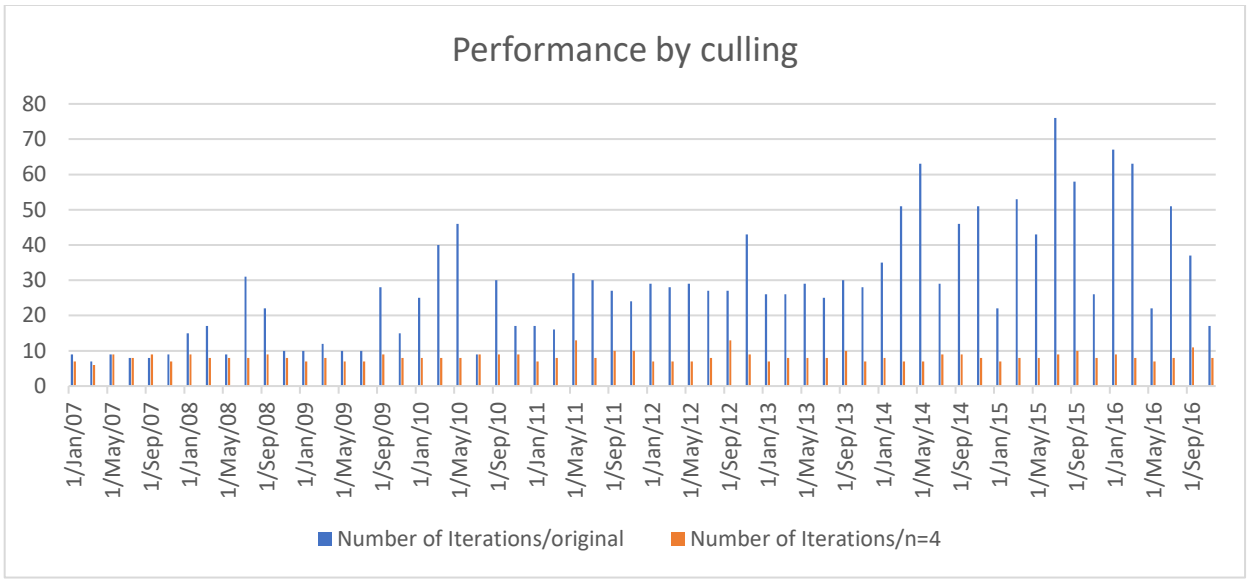



**Figure 3. 5 Performance contrast before and after being culled**

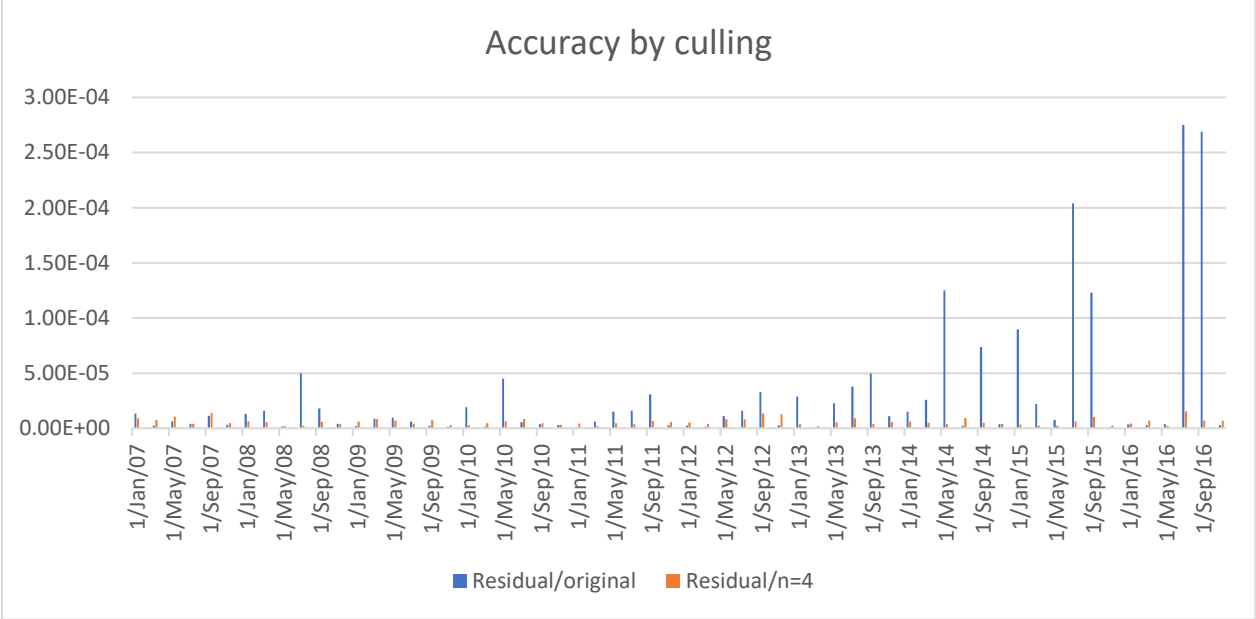



**Figure 3. 6 Result contrast before and after being culled**

**Table 3. 1 Average results of performance and accuracy with and without data culling**

| | Number of Iterations/original | Residual/original | Number of Iterations/n=4 | Residual/n=4 |
|---|---|---|---|---|
| **Average** | 28 | $5.84 \times 10^{-4}$ | 8 | $2.98 \times 10^{-5}$ |

### 3.1.2 Optimization

The purpose of the optimization strategy is adjust the control function parameters $\{\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ in order to minimize the objective function:

$$MSE_{c,y} = \frac{1}{N_k}\sum_{k=1}^{N_k}[c(y_k) - c_{y,k}]^2 \tag{3.1}$$

Where $c_{y,k}$ can be calculated from our discrete dataset, and $c(y_k)$ is the integration from $y_0$ to $y_k$:

$$c(y_k) = \int_{y_0}^{y_k} p_y dy \tag{3.2}$$

The PDF $p_y$ is transformed from $p_z$ as:

$$p_y = \frac{m}{\sigma} p_z \tag{3.3}$$

$$p_z = \exp(\int g_z \, dz) \tag{3.4}$$

$$g_z = -\left[\alpha_1 + \tan\left(\frac{\alpha_2 \pi}{180}\right) z + \sum_{n_z=1}^{N_z} \alpha_{n_z+1} z^{n_z+1}\right] \tag{3.5}$$

Minimizing the objective function is achieved by solving a non-linear least-square problem by adjusting control function parameters. Two optimization algorithms are used for this purpose: Trust Region and Levenberg–Marquardt. Both are available in Matlab and Scipy. Next, we discuss issues related to their implementation and performance in the context of this objective function.

#### *3.1.2.1 Trust-Region*

The key concept of Trust-Region method is to define a region whose radius is limited to $R_k$, or in Euclidean norm for problems with higher dimensions, around the current solution (Ye, 2014). The model is evaluated after each step. If a huge decrease is achieved, the approximate model is deemed to be successful. Otherwise, the trust region is increased provided it does not exceed the upper bound. If very subtle change in the size of the trust region occurs, then it moves forward in a new direction. In order to calculate a new step, a trust-region subproblem is solved as a quadratic model that is approximated from the objective function as:

$$\min f(x_k + p) = \min m(p) = f(x_k) + g_k^t p + \frac{1}{2} p^t B_k p \tag{3.6}$$

$$\mathrm{p} < R_k$$

where $g_k$ is the gradient at $x_k$. $B_k$ is an approximation of the real hessian matrix $H_k$ at $x_k$.
To find the minimum,

$$\frac{dm}{dp} = g_k + B_k p = 0 \quad \Leftrightarrow \quad B_k p = -g_k \tag{3.7}$$

The evaluation of the model is done by calculating

$$\rho_k = \frac{f(x_k) - f(x_k + p)}{m(0) - m(p)} \tag{3.8}$$

and is evaluated numerically.

Pseudo-code:

$x$ <-- $x_0$

while threshold is not met:

Get the improving step by solving $B_k p = -g_k$, the trust-region sub-problem

$\rho_k = \frac{f(x_k) - f(x_k + p)}{m(0) - m(p)}$

if $\rho_k < \eta_2$

$R_{k+1} <- \quad t_1 R_k$

else

if $\rho_k > \eta_3$ and p = ‖ $R_k$ ‖

$R_{k+1} <- \min(t_2 R_k, R_M)$

else

$R_{k+1} <- \quad R_k$

if $\rho_k > \eta_1$

$x_{k+1} <- \quad x_k + p_k$

else

$x_{k+1} <- \quad x_k$

*3.1.2.2 Parameter Scaling*

All of the control function parameters need to be on the scale given that the Trust-Region algorithm cannot step beyond the current trusted region. If one or a few parameters are on a much larger scale than the others, then the trust-region will have less ability to constrain the step size. This makes finding the minima of the objective function a much slower process. The control function parameters $\alpha_0,\ \alpha_1,,\ \alpha_3,\ \alpha_4$ are all dimensionless while $\alpha_2$ represents an angular slope measured in degrees with a value between $0°$ to $90°$. Initial attempts at fitting indicated that $-2 < \alpha_0,\ \alpha_1,\ \alpha_3,\ \alpha_4 < 2$ implying that they are on a different scale from $\alpha_2$. Therefore, $\alpha_2$ was replace by $\alpha_2{}' = \frac{\alpha_2}{45}$. A new starting position for the algorithm was selected as {0,0,1,0,0}.

**Table 3. 2 Average results of performance and accuracy before and after scaling**

| | Number of Iterations/Starts at [0,0,45,0,0] | Residual/Starts at [0,0,45,0,0] | Number of Iterations/Starts at [0,0,1,0,0] | Residual/Starts [0,0,1,0,0] |
|---|---|---|---|---|
| **Average** | 8.28 | $5.84 \times 10^{-6}$ | 7.27 | $4.68 \times 10^{-6}$ |

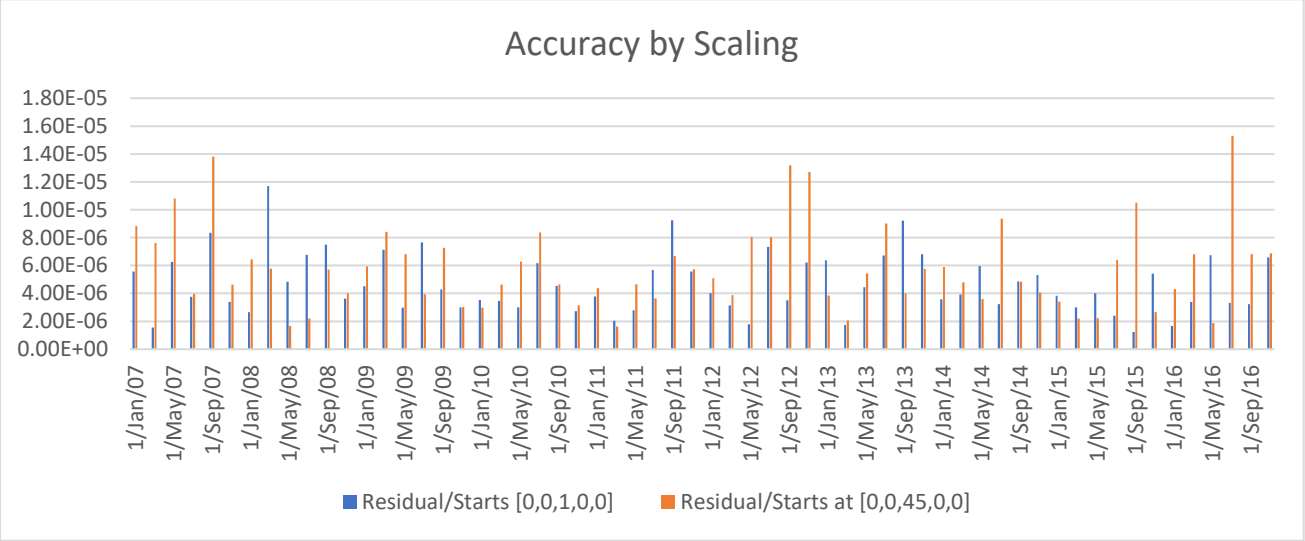



**Figure 3. 7 Results with and without scaling**

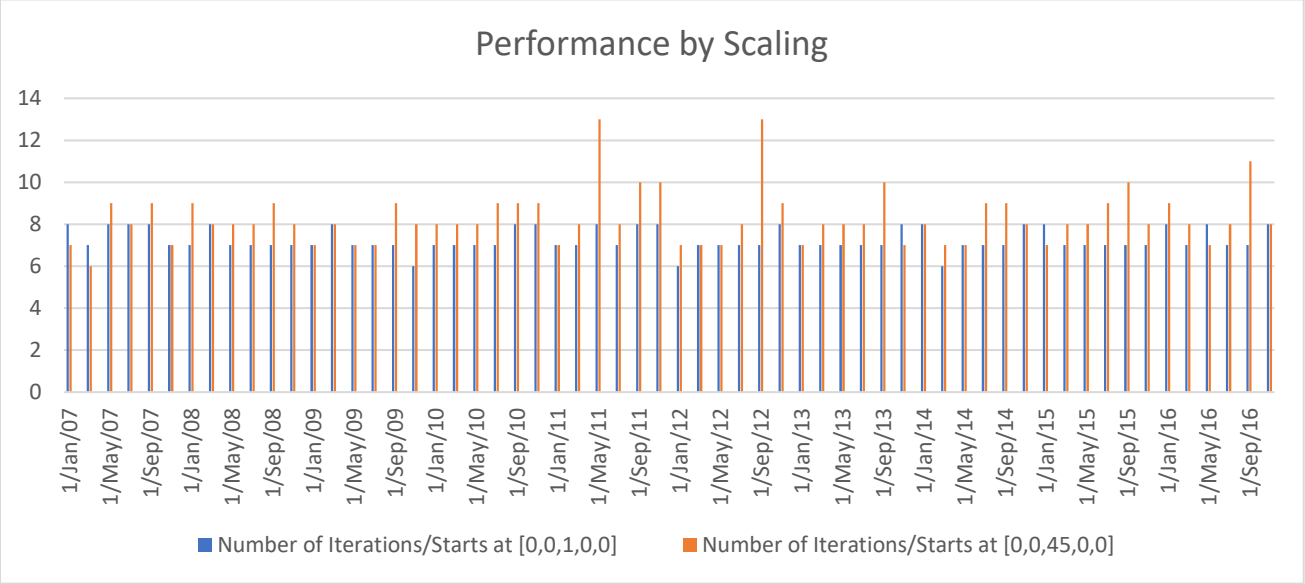



**Figure 3. 8 Performance with and without scaling**

#### *3.1.2.3 Levenberg–Marquardt algorithm*

Levernberg-Marquardt method is another traditional technique for solving non-linear least square problems. It is the combination of two other classic minimization techniques for linear regression problems: the gradient descent method, and the Gauss-Newton method. Gauss-Newton method can only be applied under the assumption that the model is quadratic. Therefore, is used when parameters are close to the local minimum. Gradient descent forces the parameter search direction to move forward in the direction steepest descent of the objective function. It is applied when parameters are further away from an optimal solution. As with the Trust-Region, the Levenberg–Marquardt algorithm also tries to solve Equation 3.7. However, Instead of setting restrictions on $p$, Levenberg-Marquardt method uses $B$ as:

$$B = H + \lambda I \tag{3.9}$$

Where $I$ is an identity matrix.

$\lambda$ is a value that gets iteratively updated.

H is the real hessian matrix of the objective function.

B is an approximation of H

When $\lambda$ is small, it is simply a Newton method. When $\lambda$ becomes large, $H$ becomes more and more negligible causing the search direction to follow the direction of steepest descent dictated by the Gradient Descent method.

#### *3.1.2.4 Comparisons*

Both the Levenberg-Marquardt and Trust Region algorithms are Newton step-based methods. The steps in their respective search directions both yield quadratic convergence behaviour for the solution variable when near the optimal solution (Berghen, F. V., 2004). The pertinent questions in this thesis is to empirically test which algorithm dominates in performance and accuracy for the water consumption data set.

In Equation 3.6, $g_k$ represents the gradient of $F(x)$ computed at $x_k$, while $p$ represents a step in the search direction. To ensure that the search direction is following the direction of steepest descent, $g_k$ and $p$ have to satisfy the condition:

$$p^t g_k < 0$$

Take the value from Equation 3.7:

$$-p^t Bp < 0 \Rightarrow p^t Bp > 0$$

Berghen (2004) provide a proof stating that if the model converges, then $B$ must be positive. The Levenberg-Marquardt algorithm ensures a positive $B$ by enlarging $\lambda$. Issues arise when $\lambda$ is an intermediate value as the algorithm then equally weighs the Newton and Gradient Descent methods in order to minimize the function. These two approach may not consistently choose the same minima and lead to poor performance of the algorithm. The direction of each step can repetitively change and slow down the convergence process (Berghen, 2004). Berghen (2004) concludes that Trust Region algorithm will thus exhibit better performances each time a negative $B$ occurs resulting in an "uphill search", and thus exhibit better performance than all the Levenberg-Marquardt algorithms. Furthermore, because Levenberg-Marquardt algorithms apply two different strategies with different efficiencies, their performance can be very sensitive to the starting point position.

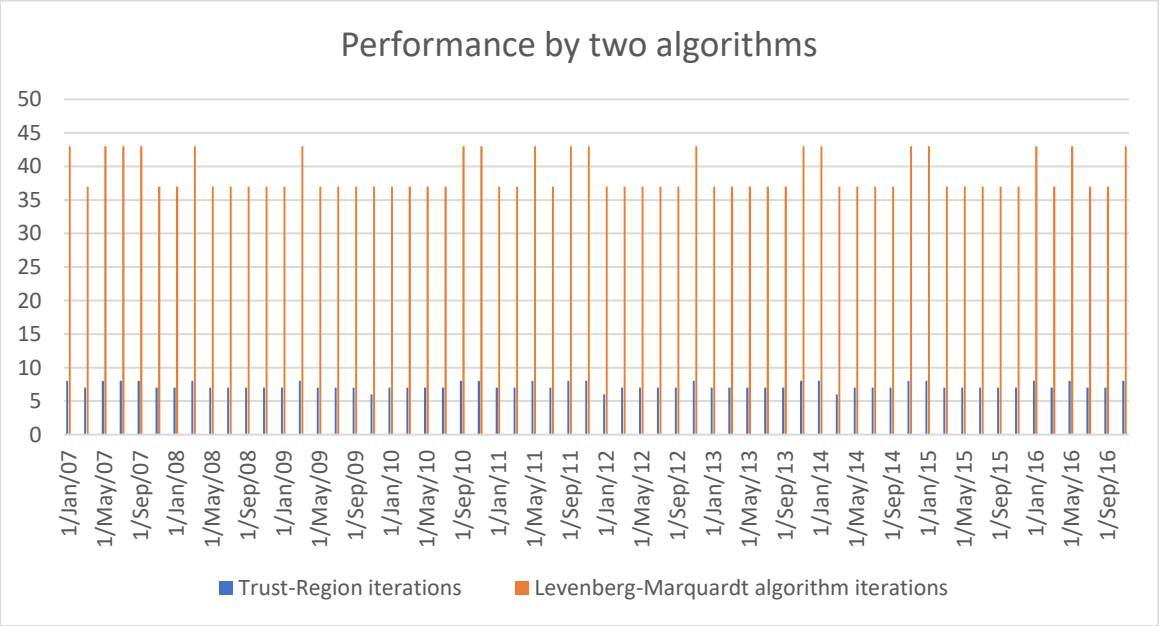



**Figure 3. 9 Trust-Region VS Levernberg-Marquardt in performace**

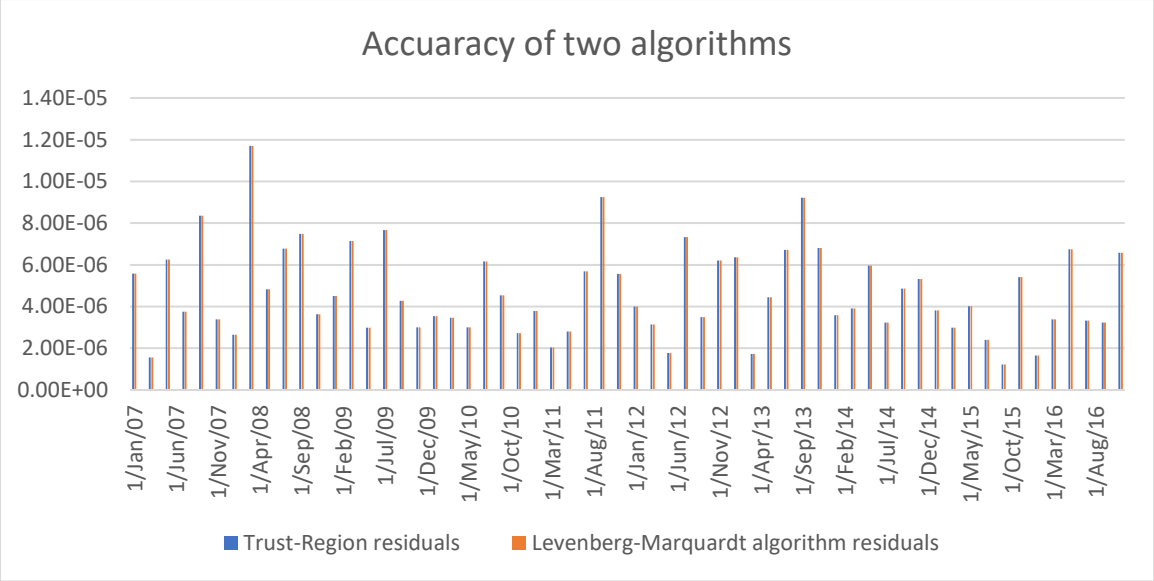



**Figure 3. 10 Trust-Region VS Levernberg-Marquardt in accuracy**

**Table 3. 3 Average results of performance and accuracy of Trust-Region and Levenberg–Marquardt**

| | Trust-Region iterations | Trust-Region residuals | Levenberg–Marquardt algorithm iterations | Levenberg–Marquardt algorithm residuals |
|---|---|---|---|---|
| Average | 7.27 | $4.68 \times 10^{-6}$ | 38.9 | $4.68 \times 10^{-6}$ |

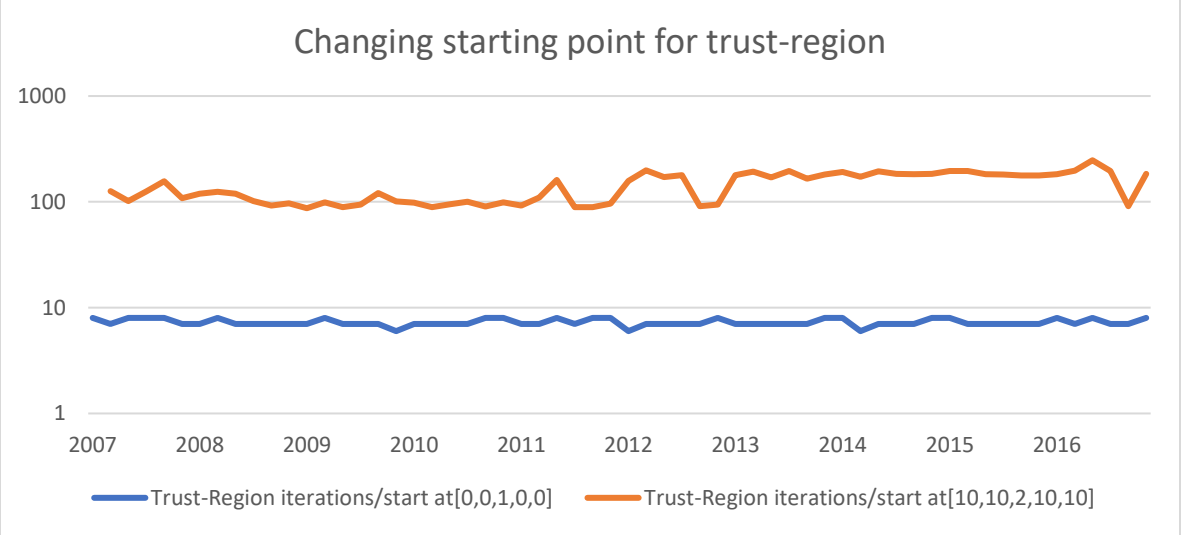



**Figure 3. 11 Trust-Region sensitivity to starting position(Performance)**

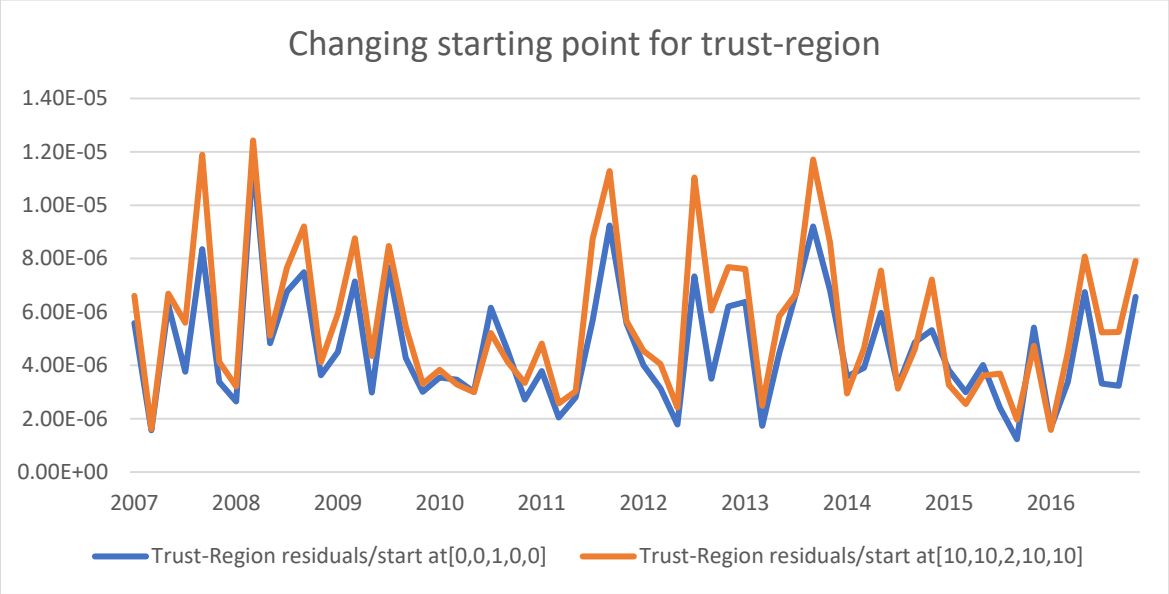



**Figure 3. 12 Trust-Region sensitivity to starting position(Accuracy)**

**Table 3. 4 Average results of performance and accuracy of Trust-Region at different starting positions**

| | Trust-Region iterations/start at[0,0,1,0,0] | Trust-Region iterations/start at[10,10,2,10,10] | Trust-Region residuals/start at[0,0,1,0,0] | Trust-Region residuals/start at[10,10,2,10,10] |
|---|---|---|---|---|
| Average | 7.27 | 142.4 | $4.68 \times 10^{-6}$ | $5.58 \times 10^{-6}$ |

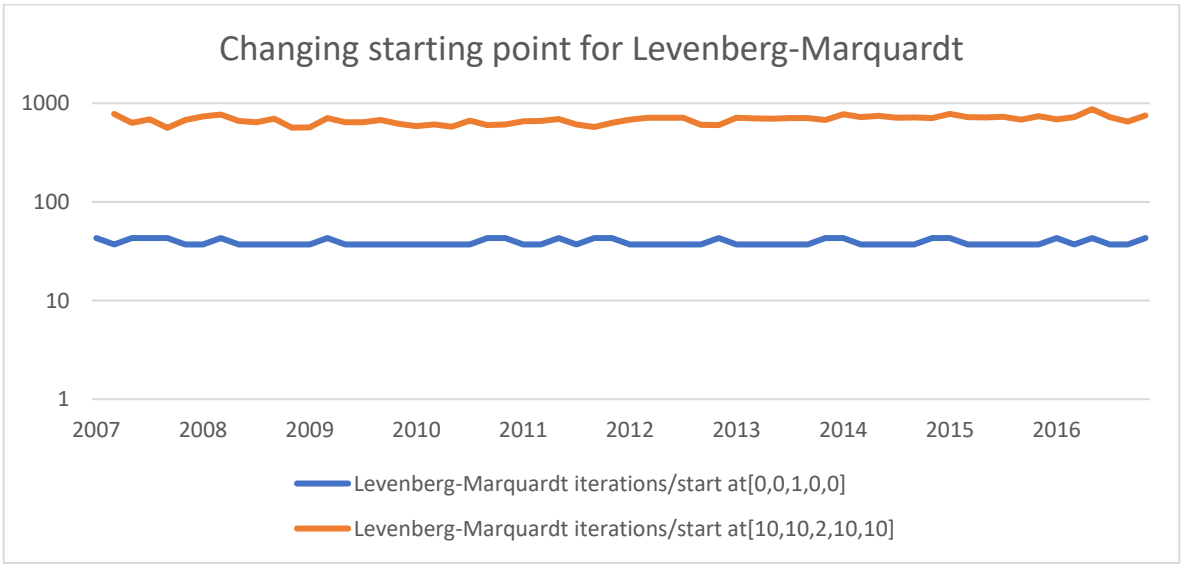



**Figure 3. 13 Levernberg-Marquardt sensitivity to starting position(Performace)**

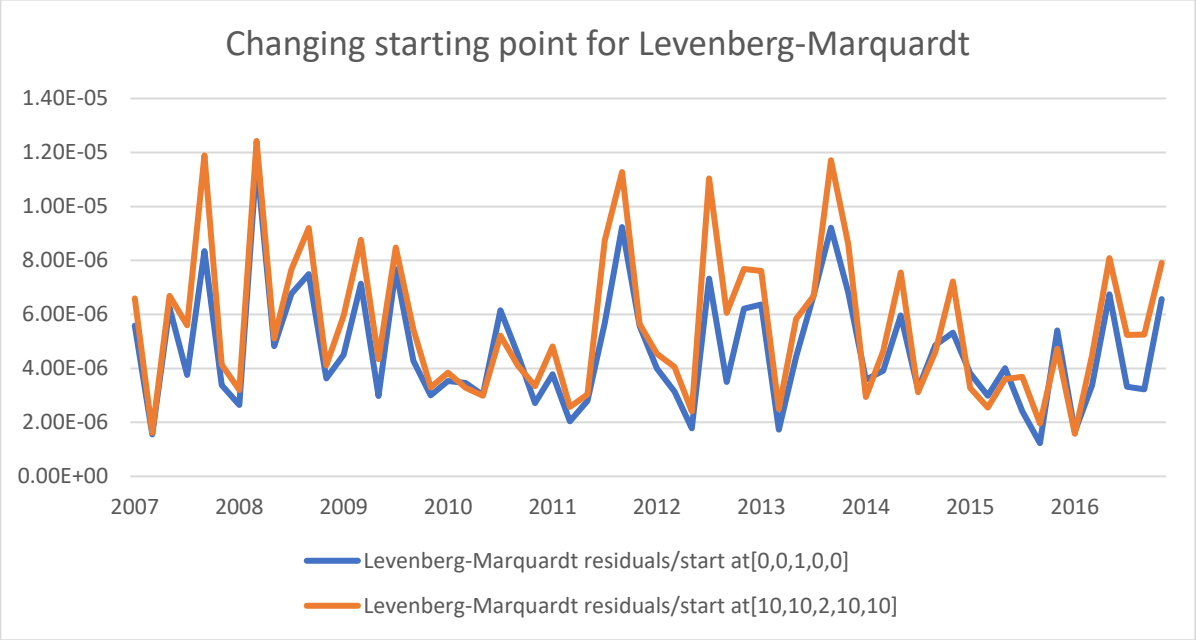



**Figure 3. 14 Levenberg-Marquardt sensitivity to starting position(Accuracy)**

**Table 3. 5 Average results of performance and accuracy of Levenberg-Marquardt at different starting positions**

| | Levenberg-Marquardt iterations/start at[0,0,1,0,0] | Levenberg-Marquardt iterations/start at[10,10,2,10,10] | Levenberg-Marquardt residuals/start at[0,0,1,0,0] | Levenberg-Marquardt residuals/start at[10,10,2,10,10] |
|---|---|---|---|---|
| Average | 7.27 | 682 | $4.68 \times 10^{-6}$ | $5.58 \times 10^{-6}$ |

In order to examine the impact of sensitivity to the starting position for both methods, we experimented with a much worse position [10,10,2,10,10]. Note that $\alpha_2{}'$ exists between 0 and 2 and was therefore set to 2. As can be seen from the Figures 3.13 and 3.14 above, choosing a bad starting position has a greater negative impact for Levenberg-Marquardt than for Trust-Region algorithm. However, both still converge to the same global minimum.

In summary, the Trust-Region method is better than Levenberg-Marquardt for both performance and accuracy, for both residential and commercial accounts. Therefore, it is used

as a default setting for the water consumption application. Because the starting position had a significant impact on performance when applying the Trust-Region, the global minima from previous solutions (at previous billing periods) are saved and then used to determine the starting solution for the next billing period.

Pseudocode:

$x_0 \leftarrow$ zeros(n)

$x_A \leftarrow$ empty array

For 1 to k:

$x$, func = least_square($fun^k$, $x_0$)

append $x$ to $x_A$

if(k = 1 or (func > 1.5*pre_func or k is multiple of 10))

$x_0$ = mean($x_A$)

### 3.1.3 Histogram Fitting Results Summary

To visualize the quality our fitting results are, we plot PMF and the PDF on the same graph for each billing period.

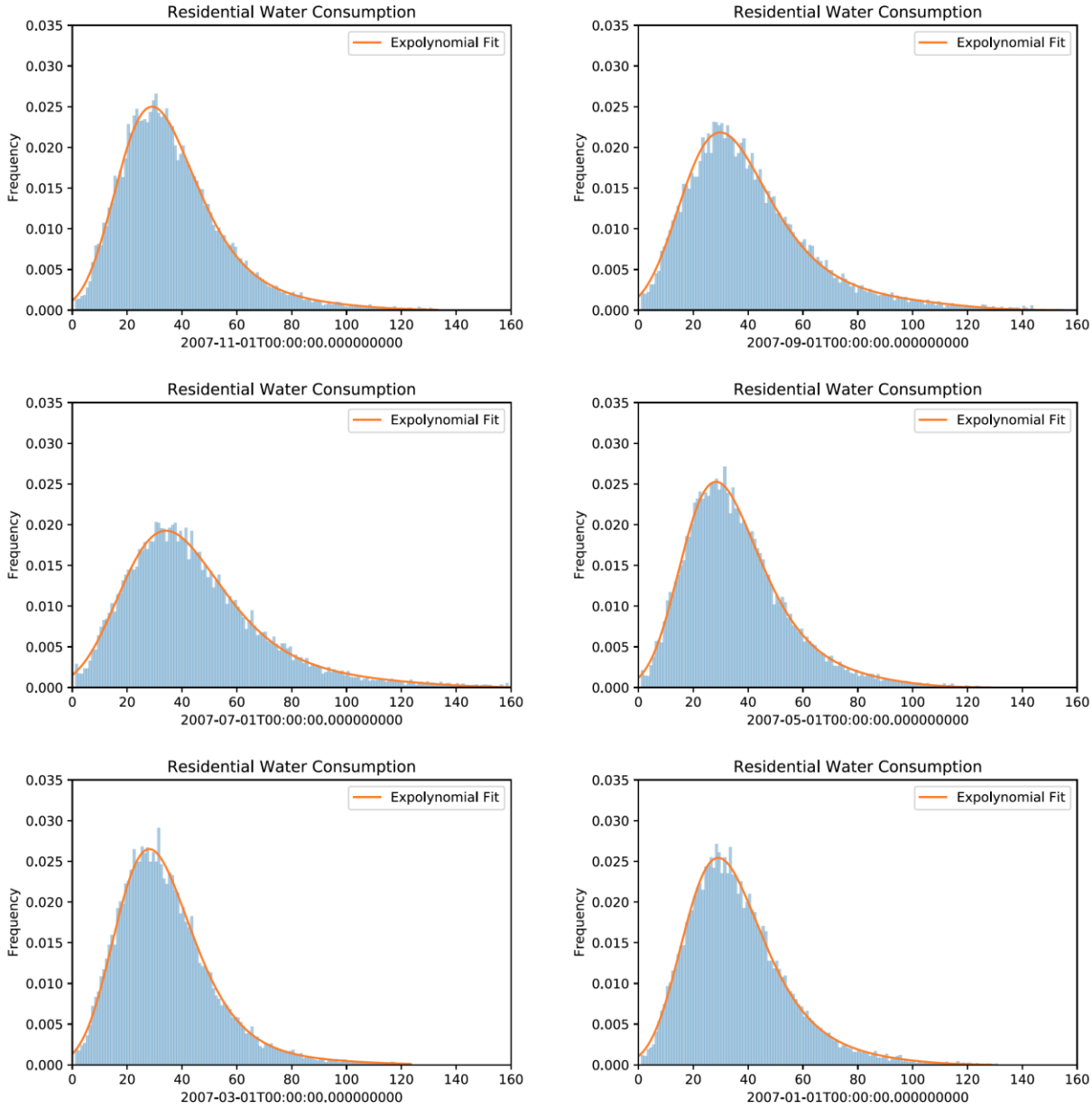



**Figure 3. 15 Residential water consumption by a single account PMF and PDF in 2007**

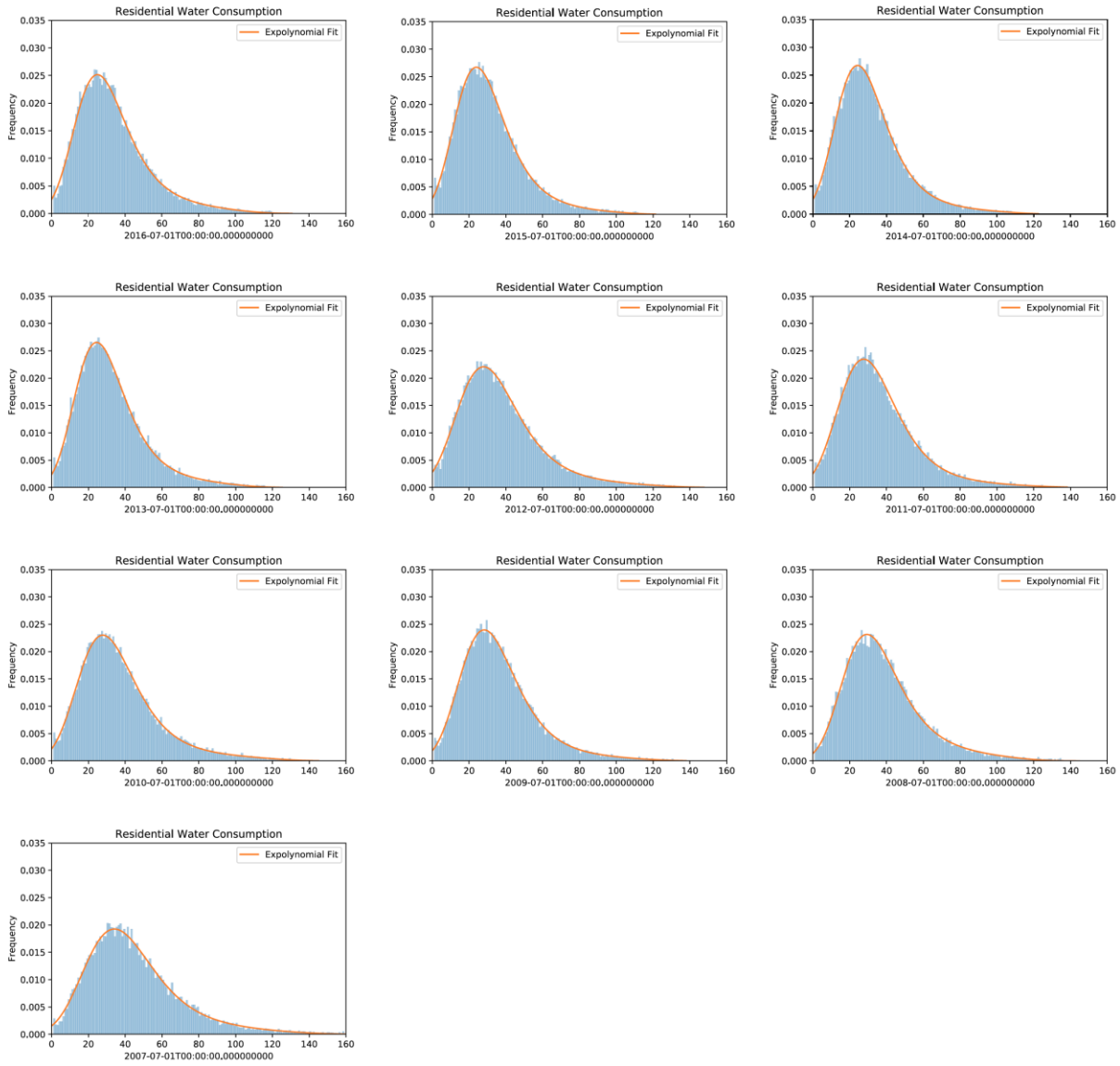



**Figure 3. 16 Residential water consumption by a single account PMF and PDF in July and August from 2007-2016**

Figure 3.16 shows residential consumption PDFs corresponding to each billing period of 2007. It demonstrates how water consumption distribution responds to changes in weather throughout the year. Closer to July and August, which are the warmest periods of time in Waterloo, the corresponding PDF shifts in a downward direction and away from the origin. When January and February approaches, it moves in the opposite direction. This meets the expectation that average consumption increases in summer, when more water is consumed and used for irrigations. In winter, outdoor water usage is limited (Enouy, 2018).

Figure 3.17 shows residential consumption PDFs of July/August every second year. The impact of increases in water price on water consumption can be seen as the corresponding PDF in each billing progressively compresses towards the origin from 2007 to 2015. This meets the expectation that average consumption drops as the price goes up. Quantifying the relationship between water demand and weather as well as price will be discussed in the next chapter.

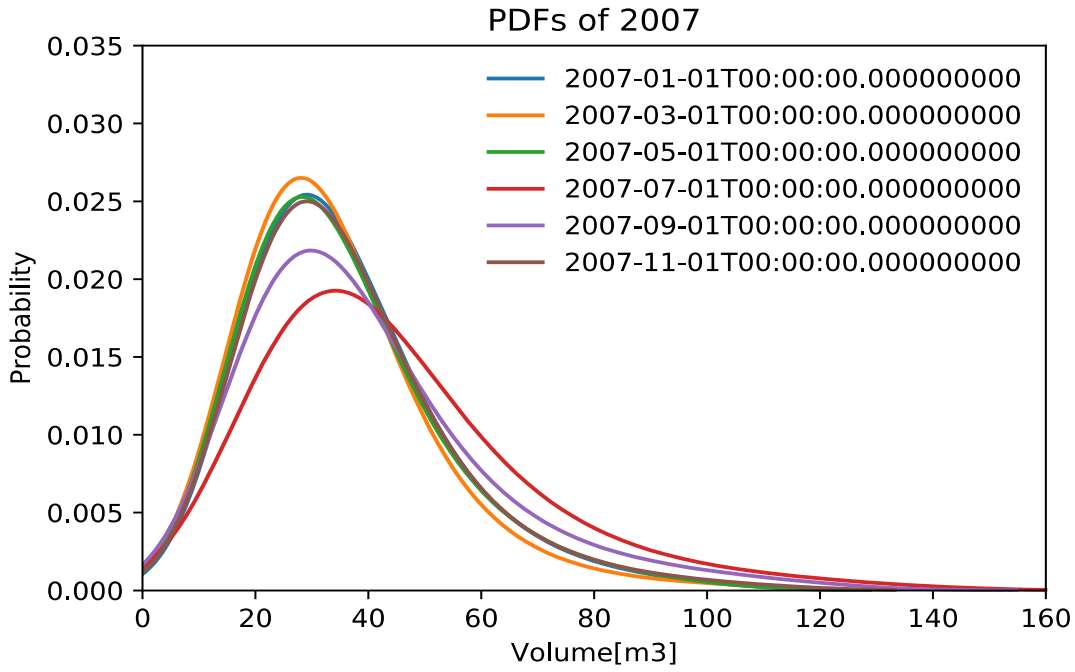



**Figure 3. 17 Residential water consumption PDFs in 2007**

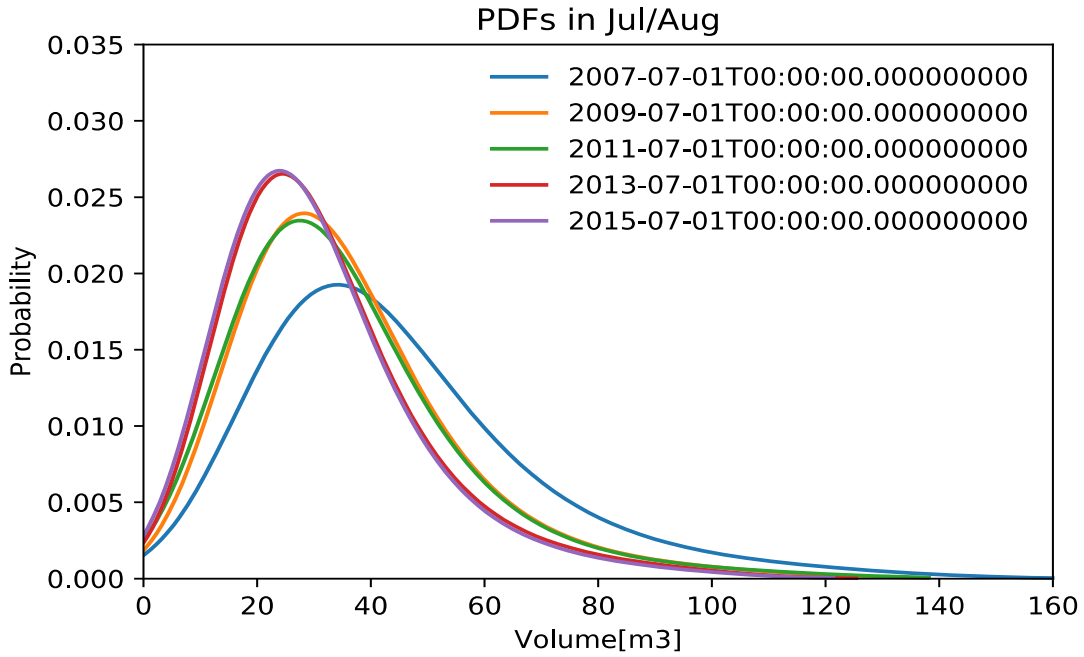



**Figure 3. 18 Residential water consumption PDFs in July and August from 2007-2016**

### 3.1.4 Regression

This section discussed whether the relationship between water consumption and price/weather should be evaluated using a linear or a curvilinear regression model as mentioned in section 2.6, and how to avoid overfitting during the parameter estimation process.

#### *3.1.4.1 Linear vs Curvilinear Regression*

When more high-order derivatives are truncated in Equation 2.17, a linear model results of the form:

$$U = b_0 + b_1 p + b_2 w \tag{3.10}$$

We begin with $\boldsymbol{U} \in \{\mu_x, \boldsymbol{m_x}, \boldsymbol{\sigma_x \alpha_0}, \boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \boldsymbol{\alpha_3}, \boldsymbol{\alpha_4}\}$. When $\boldsymbol{U} \equiv \mu_x$ we obtain the linear regression results shown on Table 3.6 indicating that linear regression in an acceptable model. Moreover, When $\boldsymbol{U} \equiv \{\boldsymbol{m_x}, \boldsymbol{\sigma_x}\}$ yields similar results.

**Table 3. 6 Linear Regression results for mean statistic**

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *R Square* | *F* |
|---|---|---|---|---|---|---|
| $b_0$ | 49.77 | 1.40 | 35.60 | $1.24 \times 10^{-40}$ | 0.77 | 96.51 ($5 \times 10^{-19}$) |
| $b_1$ | 6.69 | 0.51 | -13.12 | $7.33 \times 10^{-19}$ | | |
| $b_2$ | 0.002 | 0.0004 | 4.78 | $1.25 \times 10^{-5}$ | | |

However, we proceed to the control function parameters $\boldsymbol{U} \equiv \{\boldsymbol{\alpha_0}, \boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \boldsymbol{\alpha_3}, \boldsymbol{\alpha_4}\}$, the results indicate that a linear model is not acceptable. Table 3.7 itemizes results for $\boldsymbol{\alpha_2}$:

**Table 3. 7 Linear Regression results for $\boldsymbol{\alpha_2}$**

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *R Square* | *F* |
|---|---|---|---|---|---|---|
| $b_0$ | 1.27 | 0.029 | 43.61 | $1.78 \times 10^{-45}$ | 0.036 | 1.06 (0.35) |
| $b_1$ | -0.015 | 0.0106 | -1.41 | 0.16464766 | | |
| $b_2$ | $3 \times 10^{-6}$ | $9 \times 10^{-6}$ | 0.38 | 0.70347716 | | |

Note: F-stat
(significance of F)

The degrees of freedom in the denominator equals 2, while the degrees of freedom in the numerator is 57. This yields a $F$ critical value of 3.93 (95% confidence interval). F of $\alpha_2$ is 1.06 < 3.93, which means the null hypothesis cannot be rejected. In other words, $\alpha_2$ *might* be related to price and weather. In summary, while linear regression is acceptable for the mean, median and standard deviation statistics, it fails to explain the parametrization of the control function parameters.

Next we apply curvilinear regression $\boldsymbol{U} \in \{\mu_{\boldsymbol{x}}, \boldsymbol{m_x}, \boldsymbol{\sigma_x}\boldsymbol{\alpha_0}, \boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \boldsymbol{\alpha_3}, \boldsymbol{\alpha_4}\}$. On Table 3.8 and 3.9, regression results are itemized for $\mu_{\boldsymbol{x}}$ and $\boldsymbol{\alpha_2}$ on the basis that they failed and passed linear regression, respectively. The expectation is that curvilinear regression would explain the parametrization of both parameters. For curvilinear regression, the degrees of freedom in the denominator equals 7 while the degrees of freedom in the numerator is 52, which yields a $F$

critical value of 2.55 (95% confidence interval). $F$ of $\alpha_2$ is 2.76 < 2.55, which is sufficient to reject the null hypothesis. Curvilinear shows great improvement on our analysis, and able to quantify the relationship between $\alpha_2$ and price, weather.

**Table 3. 8 Curvilinear Regression results for mean**

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | R Square | F |
|---|---|---|---|---|---|---|
| $b_0$ | 15.62 | 12.67 | 1.23 | 0.22 | | |
| $b_1$ | 20.42 | 10.15 | 2.01 | 0.05 | | |
| $b_2$ | $1.09 \times 10^{-4}$ | 0.02 | 0.01 | 1.00 | | |
| $b_3$ | -0.01 | 0.01 | -0.42 | 0.68 | 0.84 | 38.45 |
| $b_4$ | $1.65 \times 10^{-5}$ | $7.42 \times 10^{-6}$ | 2.22 | 0.03 | | $(2.20 \times 10^{-18})$ |
| $b_5$ | $-4.89 \times 10^{-6}$ | $2.66 \times 10^{-6}$ | -1.83 | 0.07 | | |
| $b_6$ | -5.21 | 1.99 | -2.62 | 0.01 | | |
| $b_7$ | $1.92 \times 10^{-3}$ | $2.70 \times 10^{-3}$ | 0.71 | 0.48 | | |

**Table 3. 9 Curvilinear Regression results for $\alpha_2$**

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | R Square | F |
|---|---|---|---|---|---|---|
| $b_0$ | 0.80 | 0.27 | 2.95 | $4.72 \times 10^{-3}$ | | |
| $b_1$ | 0.35 | 0.22 | 1.62 | 0.11 | | |
| $b_2$ | $2.56 \times 10^{-4}$ | $3.98 \times 10^{-4}$ | 0.64 | 0.52 | | |
| $b_3$ | $-7.84 \times 10^{-5}$ | $2.98 \times 10^{-4}$ | -0.26 | 0.79 | 0.27 | 2.76 |
| $b_4$ | $-2.24 \times 10^{-7}$ | $1.59 \times 10^{-7}$ | -1.41 | 0.16 | | (0.016) |
| $b_5$ | $6.29 \times 10^{-8}$ | $5.71 \times 10^{-8}$ | 1.10 | 0.28 | | |
| $b_6$ | -0.07 | 0.04 | -1.67 | 0.10 | | |
| $b_7$ | $2.5 \times 10^{-6}$ | $5.79 \times 10^{-5}$ | 0.04 | 0.97 | | |

In the next step for curvilinear regression, we set our threshold at 20%. on the $p$-value of the $t$-statistic to evaluate the significance of a given parameter. Specifically, terms whose $p$-value is above 20% get truncated. Tables 3.10 and 3.11 tabulate curvilinear regression results for $\mu_x$ and $\boldsymbol{\alpha_2}$ following the first truncation iteration. For $\alpha_2$, the degrees of freedom in the denominator equals 3 while the degrees of freedom in the numerator is 56, which leads to an $F$ critical value of 3.34 (95% confidence interval). $F$ of $\alpha_2$ is 3.81 < 3.34, which also meets the requirement, but

did not show any improvement. We note that $p$-value of $t$-stat in $b_7$ exceeds the threshold again requiring another truncation iteration.

**Table 3.10 Curvilinear Regression results for mean terms truncated**

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | R Square | F |
|---|---|---|---|---|---|---|
| $b_0$ | 0 | N/A | N/A | N/A | | |
| $b_1$ | 31.89 | 0.79 | 40.30 | $4.73 \times 10^{-43}$ | | |
| $b_2$ | 0 | N/A | N/A | N/A | | |
| $b_3$ | 0 | N/A | N/A | N/A | 0.998 | 7444.83 |
| $b_4$ | $7.4 \times 10^{-6}$ | $2.30 \times 10^{-6}$ | 3.22 | $2.17 \times 10^{-3}$ | | $(1.82 \times 10^{-74})$ |
| $b_5$ | $-1.92 \times 10^{-6}$ | $8.35 \times 10^{-7}$ | -2.30 | 0.03 | | |
| $b_6$ | -7.31 | 0.28 | -26.00 | $6.11 \times 10^{-33}$ | | |
| $b_7$ | 0 | N/A | N/A | N/A | | |

**Table 3.11 Curvilinear Regression results for $\alpha_2$ truncate**

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | R Square | F |
|---|---|---|---|---|---|---|
| $b_0$ | 0.74 | 0.18 | 4.11 | $1.29 \times 10^{-4}$ | | |
| $b_1$ | 0.41 | 0.14 | 2.87 | 0.01 | | |
| $b_2$ | 0 | N/A | N/A | N/A | | |
| $b_3$ | 0 | N/A | N/A | N/A | 0.17 | 3.81 |
| $b_4$ | $2.06 \times 10^{-9}$ | $7.25 \times 10^{-9}$ | -0.28 | 0.78 | | (0.015) |
| $b_5$ | 0 | N/A | N/A | N/A | | |
| $b_6$ | -0.08 | 0.03 | -2.98 | $4.32 \times 10^{-3}$ | | |
| $b_7$ | 0 | N/A | N/A | N/A | | |

**Table 3.12 Curvilinear Regression results for $\alpha_2$ second truncation**

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | R Square | F |
|---|---|---|---|---|---|---|
| $b_0$ | 0.73 | 0.18 | 4.14 | $1.15 \times 10^{-4}$ | | |
| $b_1$ | 0.41 | 0.14 | 2.93 | $4.89 \times 10^{-3}$ | | |
| $b_2$ | 0 | N/A | N/A | N/A | 0.17 | 5.76 |
| $b_3$ | 0 | N/A | N/A | N/A | | (0.0053) |
| $b_4$ | 0 | N/A | N/A | N/A | | |

| | | | | |
|---|---|---|---|---|
| $b_5$ | 0 | N/A | N/A | N/A |
| $b_6$ | -0.08 | 0.03 | -3.04 | $3.56 \times 10^{-3}$ |
| $b_7$ | 0 | N/A | N/A | N/A |

After the second truncation iteration, the critical value becomes 3.15 and the model still holds. Therefore, we conclude that $\alpha_2$ does not have a relationship with weather. Instead, it is only affected by price.

In summary, curvilinear regression does significantly improve the parameter estimation process for Equation 2.17. However, iteratively truncation of all terms with $p$-values greater than 20% must be conducted to ensure that unrelated terms do not cause overfitting.

#### *3.1.4.2 Result Summary*

Results show that the mean, standard deviation and median statistics are related to both weather and water price. However, the control function parameters $\alpha_0$, $\alpha_1$, $\alpha_2$ and $\alpha_4$tend to be dependent on a single ambient process, that is: either weather or price. In contrast, $\alpha_3$ appears to be independent of both weather and price and should be discarded.

**Table 3. 13 Summary of regression results for residential water demand**

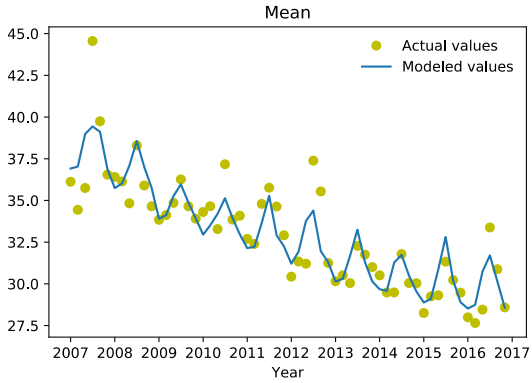
| $U$ | $b_{U,0}$ | $b_{U,1}$ | $b_{U,2}$ | $b_{U,3}$ | $b_{U,4}$ | $b_{U,5}$ | $b_{U,6}$ | $b_{U,7}$ | Regression Analysis |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\mu}_x$ | $\begin{pmatrix} 41.70 \\ 58.24 \\ 1.75 \times 10^{-52} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 1.98 \times 10^{-6} \\ 5.78 \\ 3.24 \times 10^{-7} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} -1.33 \\ -14.45 \\ 1.03 \times 10^{-20} \end{pmatrix}$ | 0.00 | † $\begin{pmatrix} 0.81 \\ 118.29 \\ 5.16 \times 10^{-21} \end{pmatrix}$ |
| $\hat{m}_x$ | $\begin{pmatrix} 25.16 \\ 3.47 \\ 1.03 \times 10^{-3} \end{pmatrix}$ | $\begin{pmatrix} 9.19 \\ 1.61 \\ 0.11 \end{pmatrix}$ | 0.00 | 0.00 | $\begin{pmatrix} 5.01 \times 10^{-6} \\ 2.61 \\ 0.01 \end{pmatrix}$ | $\begin{pmatrix} -1.22 \\ \times 10^{-6} \\ -1.75 \\ 0.08 \end{pmatrix}$ | $\begin{pmatrix} -2.87 \\ 2.6 \\ 0.01 \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 0.84 \\ 71.93 \\ 4.54 \times 10^{-21} \end{pmatrix}$ |
| $\hat{\sigma}_x$ | $\begin{pmatrix} 21.98 \\ 34.86 \\ 3.90 \times 10^{-40} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 1.77 \times 10^{-6} \\ 5.87 \\ 2.35 \times 10^{-7} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} -0.64 \\ -7.96 \\ 7.87 \times 10^{-11} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 0.62 \\ 47.32 \\ 7.75 \times 10^{-13} \end{pmatrix}$ |
| $\hat{a}_0$ | $\begin{pmatrix} 1.10 \\ 9.51 \\ 2.29 \times 10^{-13} \end{pmatrix}$ | $\begin{pmatrix} -0.28 \\ -3.05 \\ 3.47 \times 10^{-3} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 0.06 \\ 3.15 \\ 2.58 \times 10^{-3} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 0.17 \\ 5.91 \\ 4.7 \times 10^{-3} \end{pmatrix}$ |
| $\hat{a}_1$ | $\begin{pmatrix} 0.32 \\ 30.56 \\ 1.86 \times 10^{-37} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 4.38 \times 10^{-5} \\ 3.03 \\ 3.70 \times 10^{-3} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 0.14 \\ 9.15 \\ 53.7 \times 10^{-3} \end{pmatrix}$ |
| $\hat{a}_2$ | $\begin{pmatrix} 0.73 \\ 4.14 \\ 1.15 \times 10^{-4} \end{pmatrix}$ | $\begin{pmatrix} 0.41 \\ 2.93 \\ 4.89 \times 10^{-3} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} -0.08 \\ -3.04 \\ 3.56 \times 10^{-3} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 0.17 \\ 5.76 \\ 5.3 \times 10^{-3} \end{pmatrix}$ |
| $\hat{a}_3$ | $\begin{pmatrix} -0.71 \\ -28.29 \\ 1.25 \times 10^{-35} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} -7.37 \times 10^{-5} \\ -2.12 \\ 0.04 \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 0.07 \\ 4.5 \\ 3.8 \times 10^{-2} \end{pmatrix}$ |
| $\hat{a}_4$ | $\begin{pmatrix} 0.10 \\ 16.65 \\ 9.22 \times 10^{-24} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 2.07 \times 10^{-5} \\ 2.44 \\ 0.02 \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 0.09 \\ 5.94 \\ 1.7 \times 10^{-2} \end{pmatrix}$ |

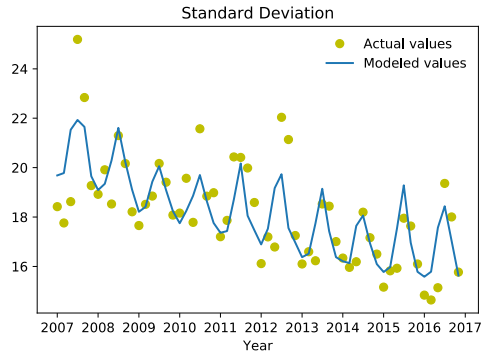
Notes: Total degrees of freedom is 59 and residual degrees of freedom is 57.

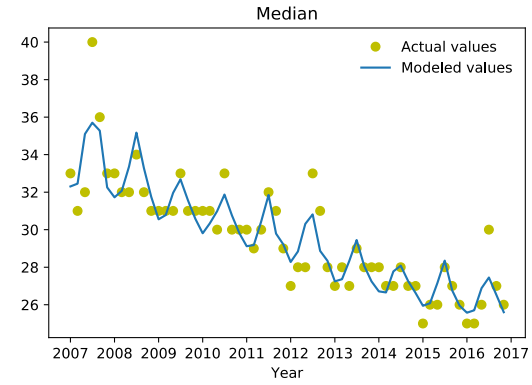
† $\begin{pmatrix} R^2 \\ F \text{ statistic} \\ p-\text{value for } F \text{ statistic} \end{pmatrix}$

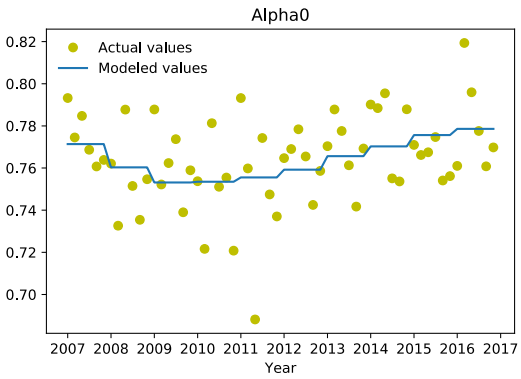
‡ $\begin{pmatrix} \text{Regression Coefficient} \\ t \text{ statistic} \\ p-\text{value for } t \text{ statistic} \end{pmatrix}$

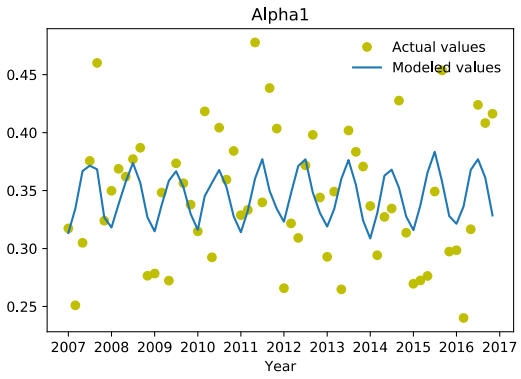
Figure 3.19 is the visualization of our regression results.



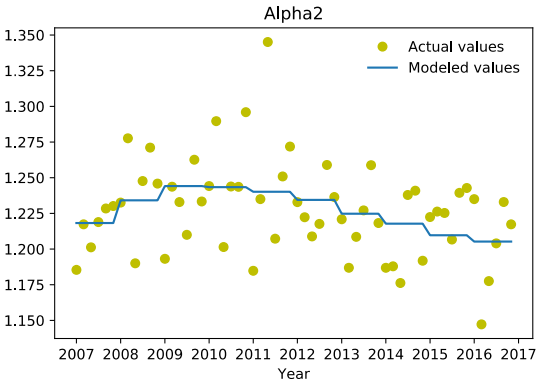




**Figure 3. 19 Modeled Results vs Actual Results for residential water demand**

## 3.2 Commercial Water Demand Analysis

In contrast to the large number of residential accounts, the city only has had 2957 commercial accounts created in the past ten years, with only approximately 1000 accounts active in each billing period. This section provides an analysis of water consumption using these records.

### 3.2.1 Lognormal of the Median-Relative

None of residential water consumption histogram data fall under the category of heavy-tailed distribution. However, all of the commercial water consumption data are classified as heavy-tailed distributions. Even when culling ratio $y_{max} = 10$, more than 10% of the data still get removed. Therefore, we decide to fit the data using the logarithm of the median-relative space which we denote it as $\psi$. This transformation satisfies the following properties:

$$\int p_x^* \, dx = \int p_\psi^* \, d\psi = \int p_z \, dz \tag{3.11}$$

Furthermore, Table 3.13 summarizes the transformations for continuous zero-centered PDFs between each spatial representation.

**Table 3. 14 Spatial transformation in measurement, logarithm median-relative and standard-score space**

| Space | Magnitude | PDF | Derivative |
|---|---|---|---|
| $\boldsymbol{x}$ | $x_i$ | $p_x^* = \frac{1}{\sigma_{x,i}} p_z$ | $dx = \sigma_{x,i} dz$ |
| $\psi$ | $\psi_i = \ln\left(\frac{x_i}{m_{x,i}}\right)$ | $p_\psi^* = \frac{1}{\ln\left(\frac{x_i}{m_{x,i}}\right)} p_z$ | $d\psi = \ln\left(\frac{x_i}{m_{x,i}}\right) dz$ |
| $\mathbf{z}$ | $z_i = \frac{\psi_i}{\sigma_{\ln\left(\frac{x_i}{m_{x,i}}\right)}}$ | $p_z$ | $dz = \frac{1}{\ln\left(\frac{x_i}{m_{x,i}}\right)} d\psi$ |

Figure 3.20 depicts the City of Waterloo commercial water consumption data from Jan/Feb 2008 before and after being transformed to $\psi$ space.

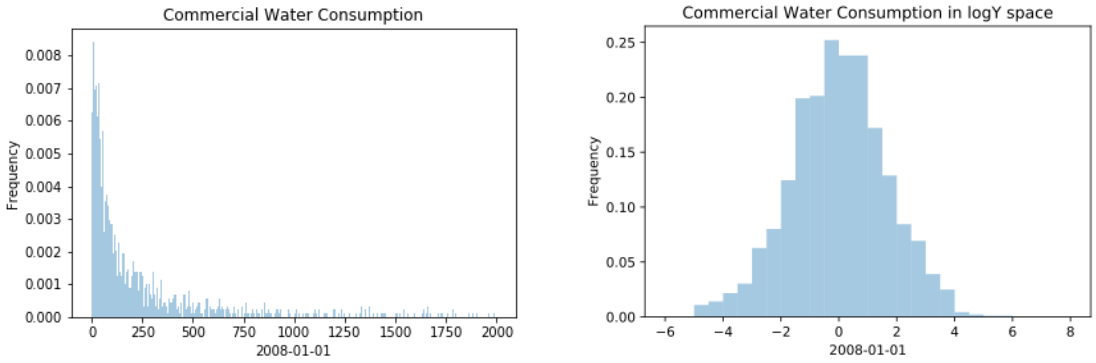



**Figure 3.20 Commercial water histogram before and after transformation log(y) space**

Table 3.14 shows the result of directly applying expolynomial series on the whole dataset and compared to that after being transformed. This transformation significantly improved both performance and the accuracy of the final result.

**Table 3.15 Average results of performance and accuracy of fitting in y and logY space**

| | LogY Space/Iterations | LogY Space/Residuals | Y Space/Iterations | Y Space/Residuals |
|---|---|---|---|---|
| Average | 8.87 | $2.66 \times 10^{-5}$ | 41.9 | 0.0047 |

### 3.2.2 Result Summary

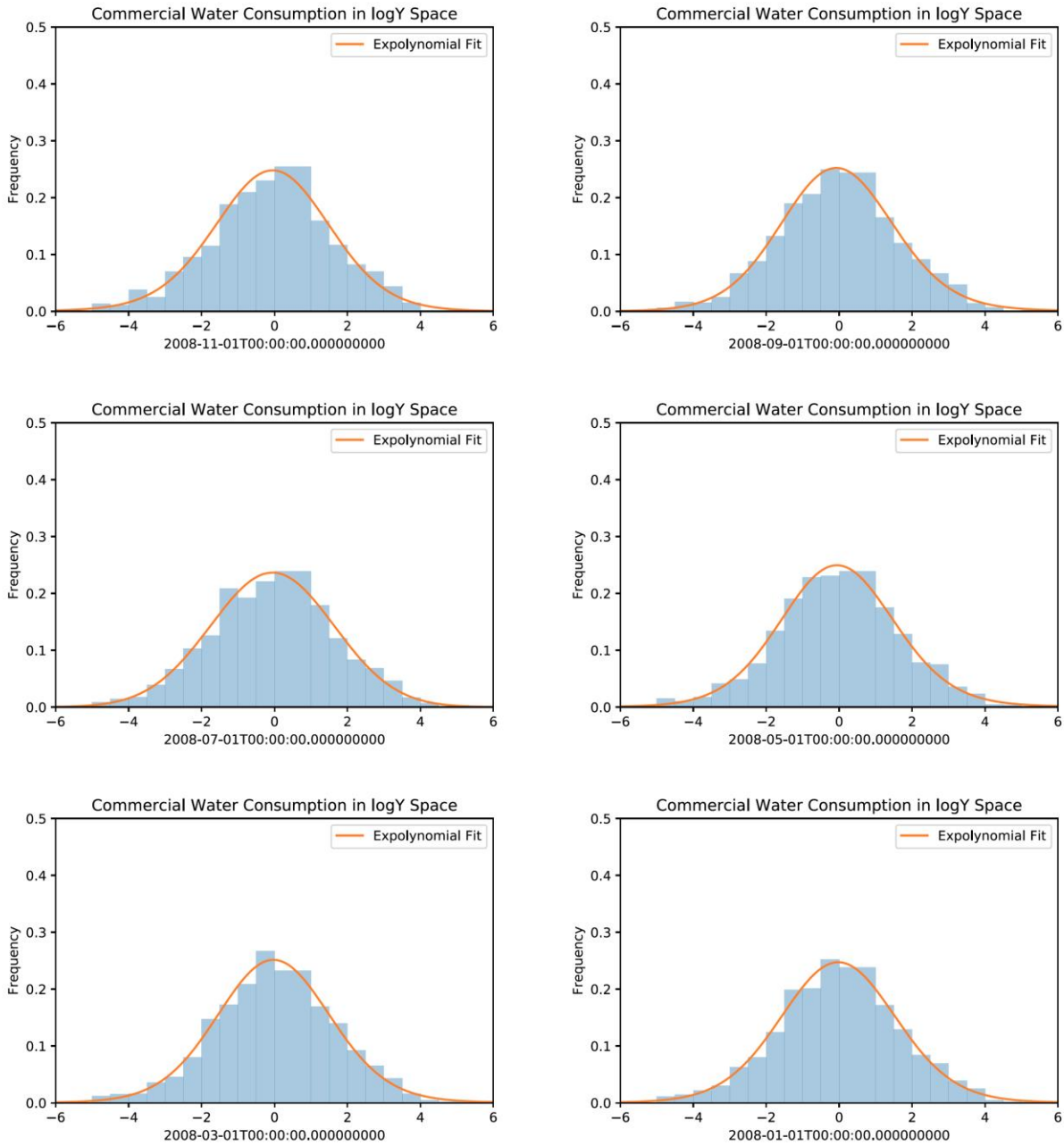



**Figure 3. 21 Commercial water consumption by a single account PMF and PDF in 2007**

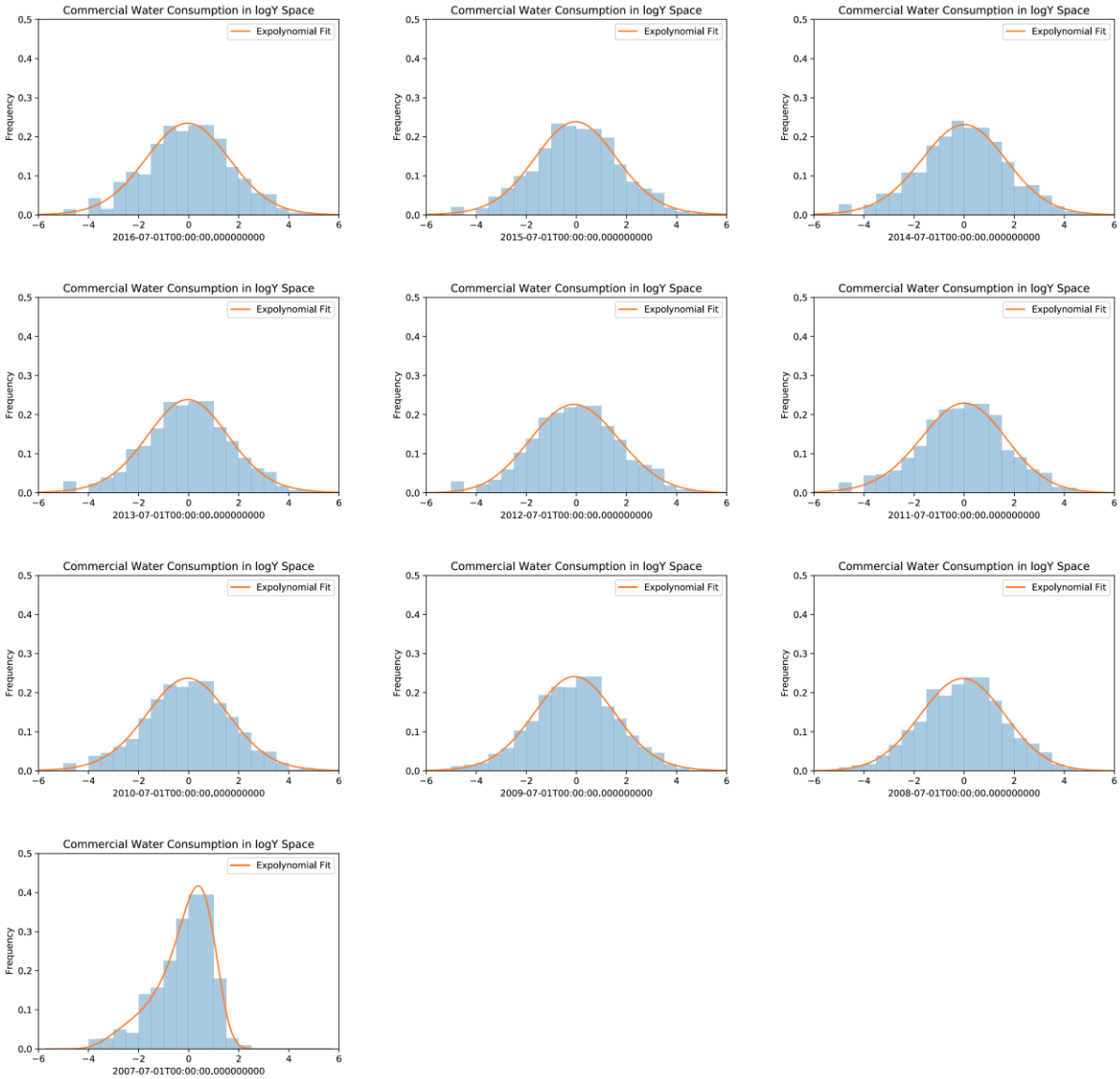



**Figure 3. 22 Commercial water consumption by a single account PMF and PDF in July and August from 2007-2016**

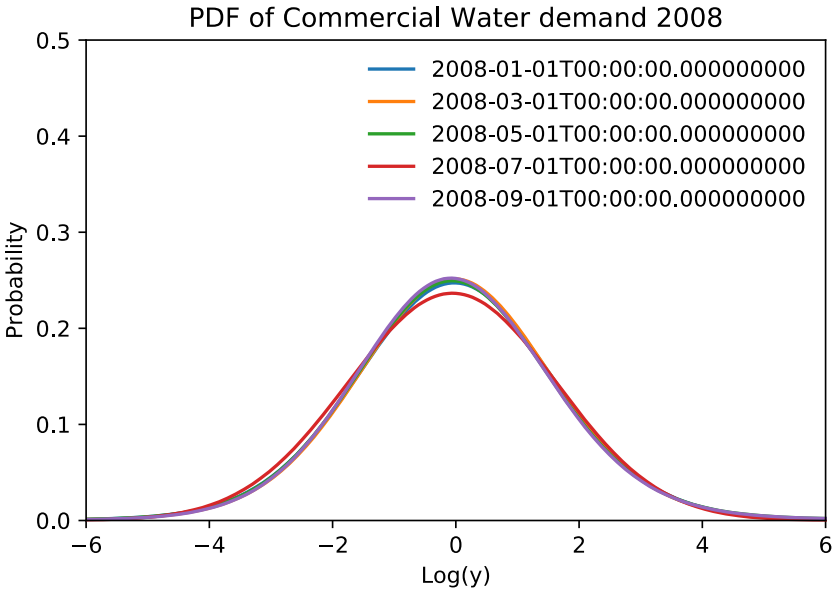



**Figure 16 Commercial water consumption by a single account PDFs in 2007**

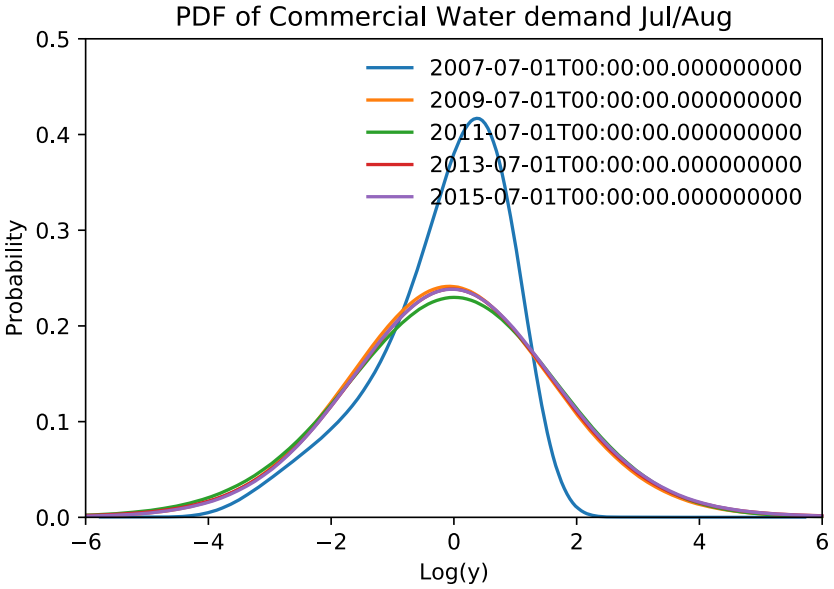



**Figure 3. 23 Commercial water consumption by a single account PDFs in July and August from 2007-2016**

The fitted PDF from 2007 deviates significantly from those in subsequent years. Water consumption by commercial accounts in 2007 is lower than in later periods. This could be a result of a change in the manner in which accounts were classified. For instance, some of the account could have been reclassified as residential accounts in later years. Therefore, results from 2007 were not used in this regression.

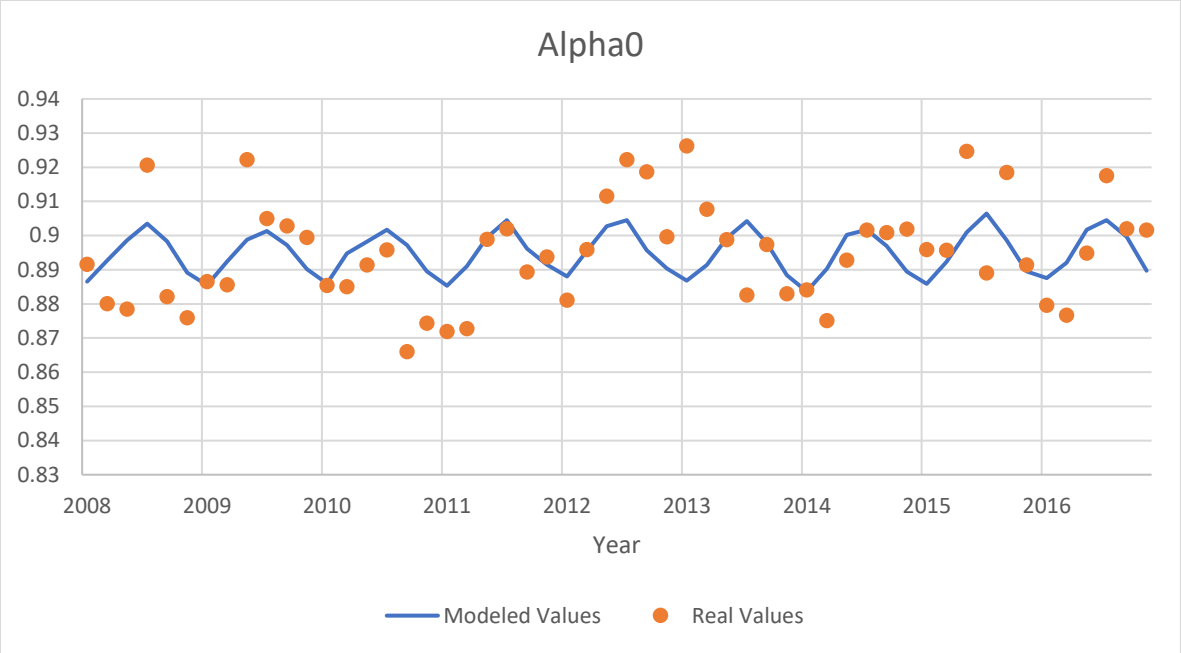
**Table 3. 16 Summary of regression results for commercial water demand**

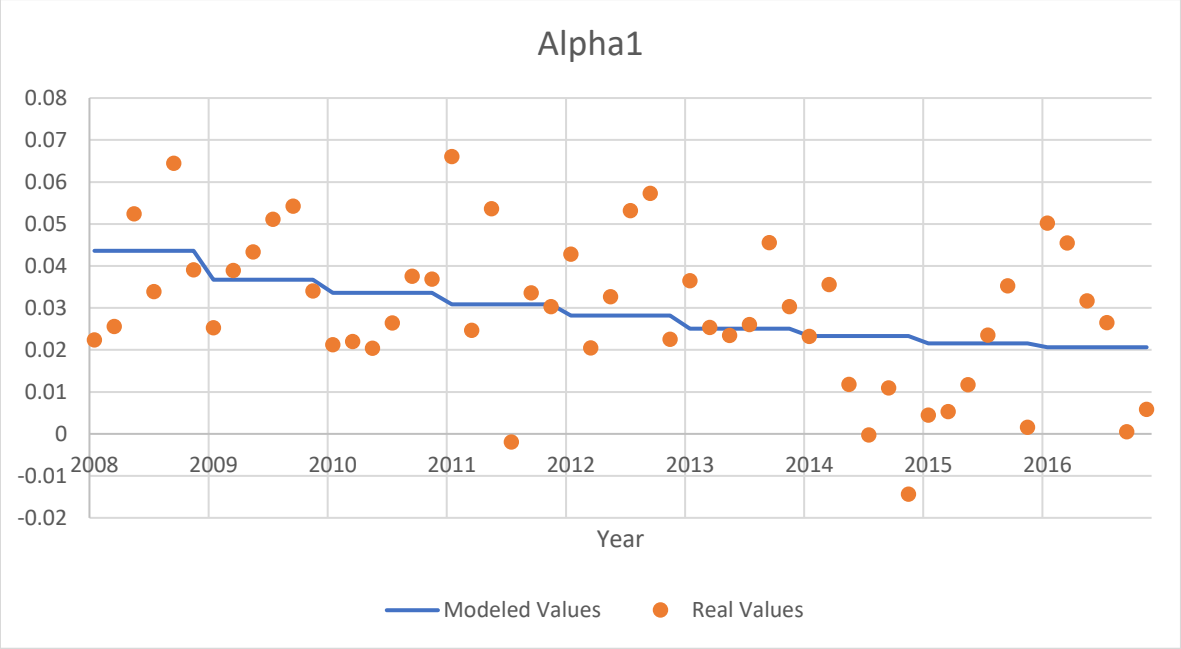
| $U$ | $b_{U,0}$ | $b_{U,1}$ | $b_{U,2}$ | $b_{U,3}$ | $b_{U,4}$ | $b_{U,5}$ | $b_{U,6}$ | $b_{U,7}$ | Regression Analysis |
|---|---|---|---|---|---|---|---|---|---|
| $\hat{\mu}_x$ | $\begin{pmatrix} 648.98 \\ 18.05 \\ 8.51 \times 10^{-24} \end{pmatrix}$ | $\begin{pmatrix} -67.12 \\ -5.21 \\ 3.38 \times 10^{-6} \end{pmatrix}$ | $\begin{pmatrix} 1.20 \times 10^{-2} \\ -2.21 \\ 3.19 \times 10^{-2} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $\dagger \begin{pmatrix} 0.39 \\ 16.49 \\ 2.99 \times 10^{-6} \end{pmatrix}$ |
| $\hat{m}_x$ | $\begin{pmatrix} 291.98 \\ 7.49 \\ 1.03 \times 10^{-9} \end{pmatrix}$ | $\begin{pmatrix} -116.50 \\ -3.94 \\ 2.50 \times 10^{-4} \end{pmatrix}$ | 0.00 | 0.00 | $\begin{pmatrix} 2.85 \times 10^{-6} \\ -2.77 \\ 7.89 \times 10^{-3} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 17.42 \\ 3.15 \\ 2.77 \times 10^{-3} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 0.81 \\ 70.03 \\ 6.46 \times 10^{-18} \end{pmatrix}$ |
| $\hat{\sigma}_{xlog}$ | $\begin{pmatrix} 0.98 \\ 4.84 \\ 1.23 \times 10^{-5} \end{pmatrix}$ | $\begin{pmatrix} 0.49 \\ 3.18 \\ 2.52 \times 10^{-3} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} -0.078 \\ -2.74 \\ 8.48 \times 10^{-3} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 0.56 \\ 32.46 \\ 8.09 \times 10^{-10} \end{pmatrix}$ |
| $\hat{a}_0$ | $\begin{pmatrix} 0.89 \\ 304.15 \\ 3.40 \times 10^{-86} \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 1.36 \times 10^{-5} \\ 3.29 \\ 1.78 \times 10^{-3} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 0.17 \\ 10.85 \\ 1.78 \times 10^{-3} \end{pmatrix}$ |
| $\hat{a}_1$ | $\begin{pmatrix} 0.09 \\ 4.78 \\ 1.50 \times 10^{-5} \end{pmatrix}$ | $\begin{pmatrix} -0.022 \\ -3.26 \\ 1.97 \times 10^{-3} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 0.17 \\ 10.63 \\ 1.97 \times 10^{-3} \end{pmatrix}$ |
| $\hat{a}_2$ | $\begin{pmatrix} 1.15 \\ 30.81 \\ 1.26 \times 10^{-34} \end{pmatrix}$ | $\begin{pmatrix} -0.023 \\ -1.74 \\ 8.84 \times 10^{-2} \end{pmatrix}$ | $\begin{pmatrix} -3.10 \\ \times 10^{-5} \\ -3.32 \\ 1.69 \times 10^{-3} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 0.22 \\ 7.24 \\ 1.71 \times 10^{-3} \end{pmatrix}$ |
| $\hat{a}_3$ | $\begin{pmatrix} -0.030 \\ -1.93 \\ 6.0 \times 10^{-2} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 0.0031 \\ 1.59 \\ 0.12 \end{pmatrix}$ | 0.00 | $\begin{pmatrix} 0.047 \\ 2.54 \\ 0.12 \end{pmatrix}$ |
| $\hat{a}_4$ | $\begin{pmatrix} -0.09 \\ -4.22 \\ 9.88 \times 10^{-5} \end{pmatrix}$ | $\begin{pmatrix} 0.015 \\ 1.96 \\ 5.53 \times 10^{-2} \end{pmatrix}$ | $\begin{pmatrix} 1.69 \times 10^{-5} \\ 3.15 \\ 2.74 \times 10^{-3} \end{pmatrix}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | $\begin{pmatrix} 0.22 \\ 7.13 \\ 1.86 \times 10^{-3} \end{pmatrix}$ |

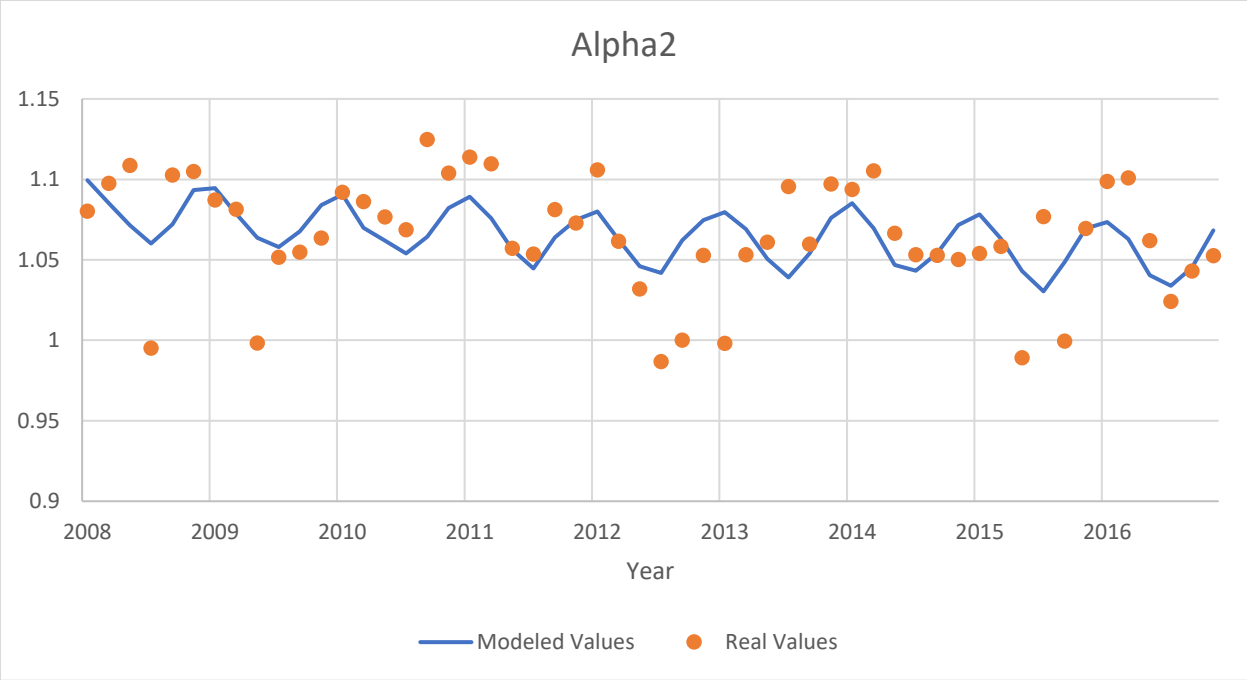
Notes: Total degrees of freedom is 53 and residual degrees of freedom is 51.

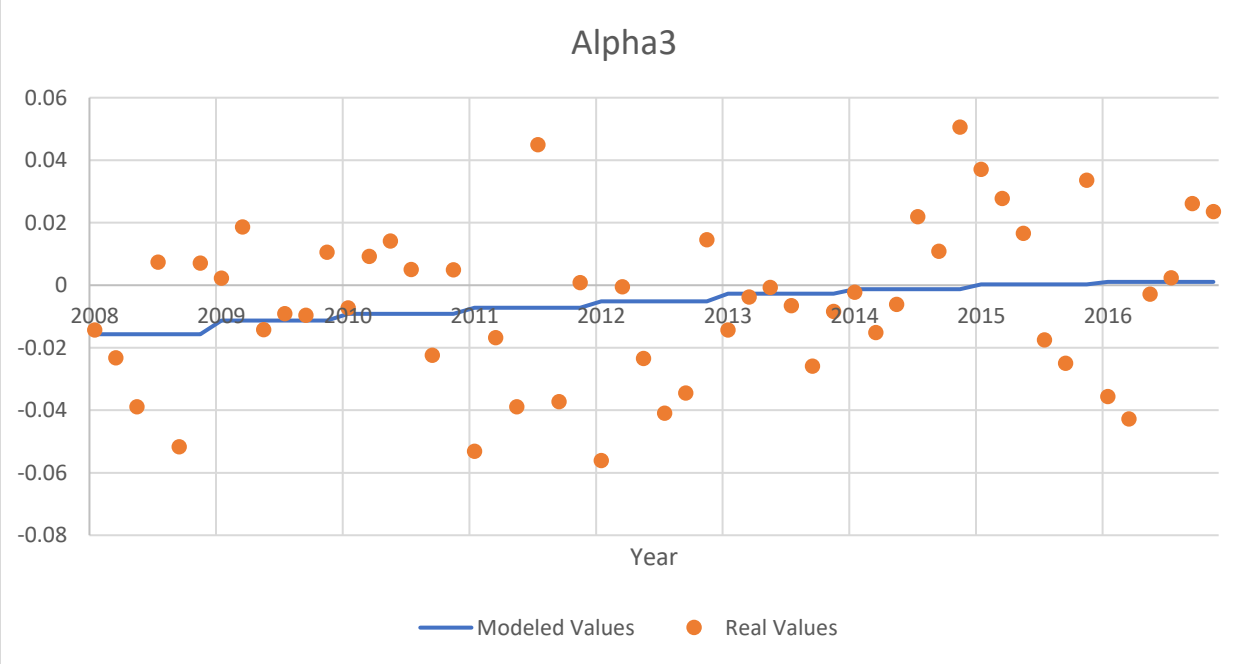
$\dagger \begin{pmatrix} R^2 \\ F \text{ statistic} \\ p-\text{value for } F \text{ statistic} \end{pmatrix}$

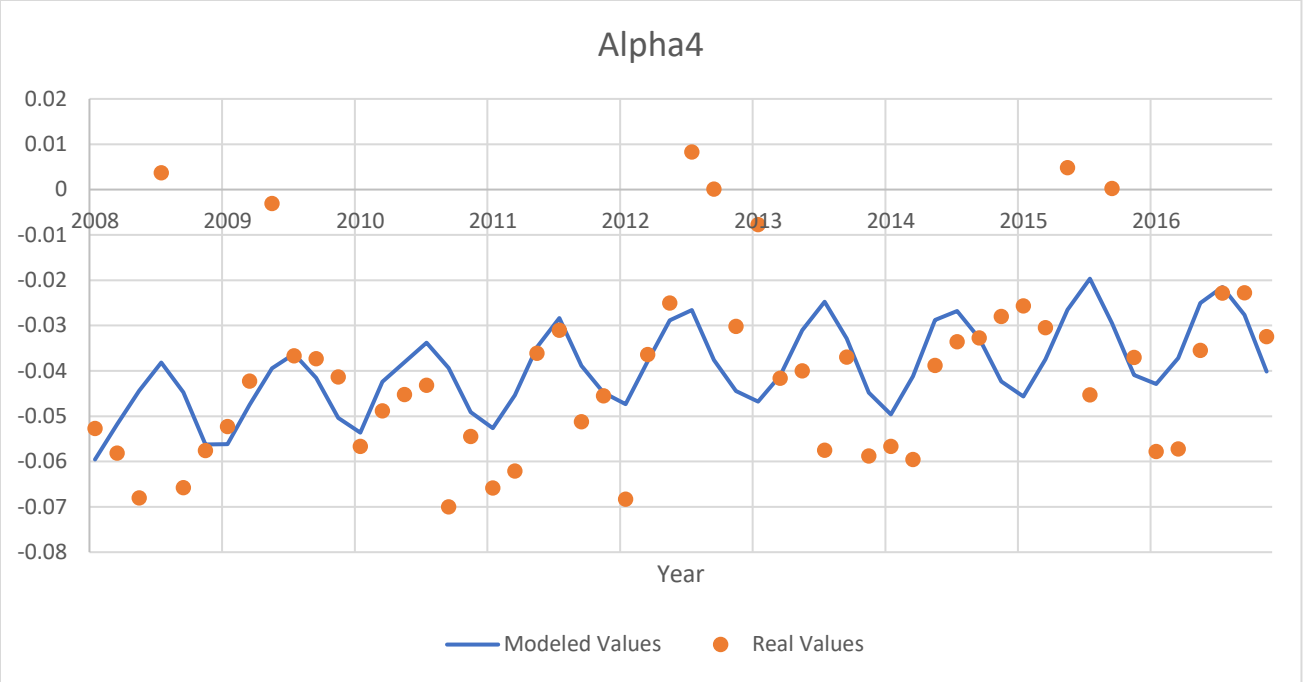
$\ddagger \begin{pmatrix} \text{Regression Coefficient} \\ t \text{ statistic} \\ p-\text{value for } t \text{ statistic} \end{pmatrix}$

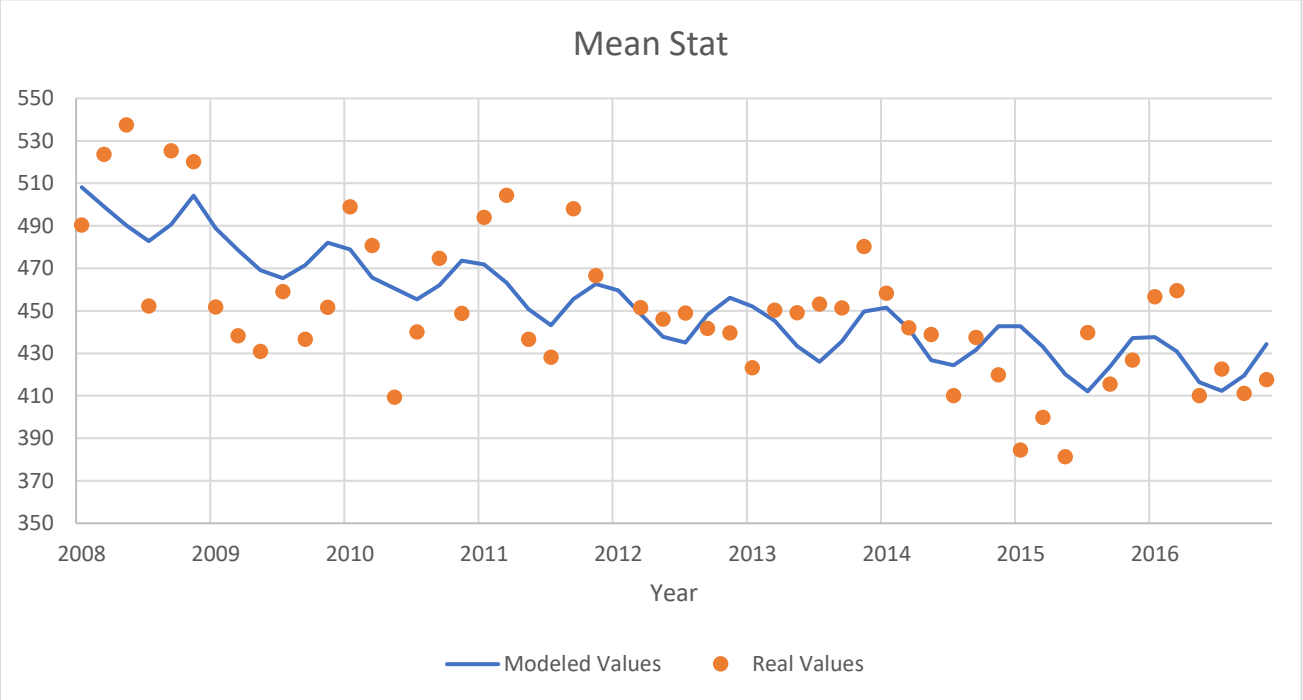

Alpha0
0.94
0.93
0.92
0.91
0.9
0.89
0.88
0.87
0.86
0.85
0.84
0.83
2008
2009
2010
2011
2012
2013
2014
2015
2016
Year
Modeled Values
Real Values

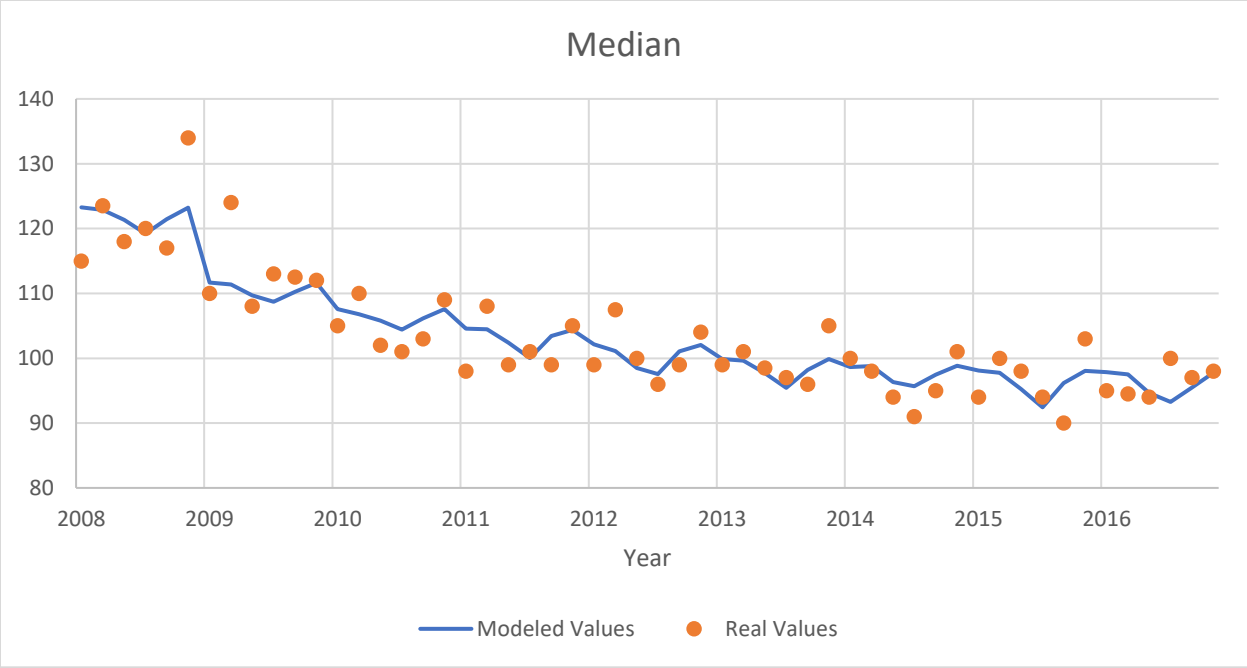

Alpha1
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
2008
2009
2010
2011
2012
2013
2014
2015
2016
Year
Modeled Values
Real Values


Alpha2
1.15
1.1
1.05
1
0.95
0.9
2008
2009
2010
2011
2012
2013
2014
2015
2016
Year
Modeled Values
Real Values


Alpha3
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
2008
2009
2010
2011
2012
2013
2014
2015
2016
Year
Modeled Values
Real Values


Alpha4
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
-0.06
-0.07
-0.08
2008
2009
2010
2011
2012
2013
2014
2015
2016
Year
Modeled Values
Real Values


Mean Stat
550
530
510
490
470
450
430
410
390
370
350
2008
2009
2010
2011
2012
2013
2014
2015
2016
Year
Modeled Values
Real Values



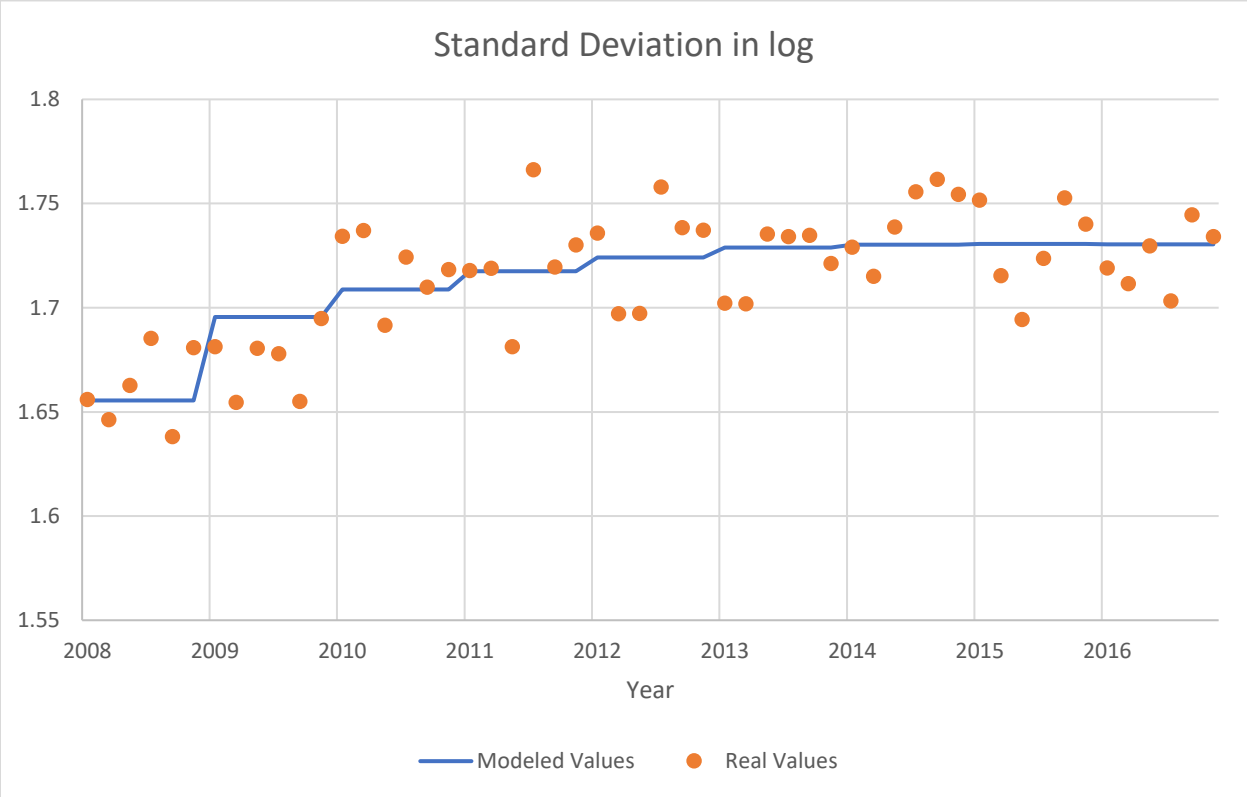




**Figure 3. 24 Model Results vs Real Results for commercial water demand**

# 4 SOFTWARE BUILDING WORKFLOW

Unlike most software programs which focus on directing traffic flow, the design elements of a numerical or data analysis software are in its algorithm. The complexity of the algorithm determines how difficult it is for a software engineer to understand and specify the requirements. With reference to this analysis, the following issues were considered.

1) Errors in original requirements

The original requirements were presented in the form of a paper. Four examples were initially implemented in am Excel spreadsheet. These implementations were then broken down into a few steps, with some steps implemented manually. At this point, implicit errors could have been made by the client during mathematical processes like constructing numerical integrals and derivatives. These errors might only result in incorrect results for some specific cases that haven't been implemented and examined in Excel as test problems. Being able to discover these errors requires the software engineer to have a solid math background and to fully understand the algorithm.

2) Uncertainties in results

The program can be divided into two parts: data compression, and regression analysis. Data compression involves fitting the histogram with a continuous probability density function and then representing the discrete data as its mean, median, standard deviation and function parameters. These statistics are a product of fitting, with the final result being dependent upon both visual inspection and by minimizing a residual value of an error function. Constrains on the residual value are a function of the accuracy of the various solvers implemented in the process. Finally, the data obtained from the city of Waterloo only covers the last ten years of water consumption. Thus may not be a sufficiently large enough data sample to determine whether the multivariate curvilinear regression analysis can successfully determine the dependence of the water consumption on ambient process of price and weather.

3) Testing of results

Constructing tests is a difficult process given the uncertainties in the results themselves. In many cases, the data compression results can be visually examined for accuracy. However, poor fitting results do not necessarily indicate an implementation failure. Instead, they could arise from limitations of algorithm when fitting certain data sets. Occasionally, ideal fitting results can also arise from a poor implementation of the algorithm itself.

4) Performance attributes

The original algorithm specified by client was implement in Excel. Therefore, it is difficult to benchmark the performance of this algorithm given the limitations in the original design. Some metrics are reasonable such as number of non-linear solver iterations to achieve the minima of the objective function.

## 4.1 Database Construction

The current database contains more than ten million records consisting of observations from the cities of Waterloo and London, Ontario. The intent is to be able to scale the database to include even more records. Hence, it is imperative to promote efficiency as part of this future expansion. To promote efficacy, database indexes are used on various items to reduce run time from $\mathcal{O}(n)$ to $\mathcal{O}(\log n)$. Future improvements can also be achieved by improving logical plans or query optimizations.

The water demand database is designed and constructed in MySQL. There are three entities: **Account**, **Consumption**, and **Location**. Furthermore, there are two sub-entities of **Account**: **Municipal** and **Res_type**. **Account** consists all the accounts from the different cities. **Account_No** is the primary key representing local account numbers, and is further qualified by appending **Municipal_id** to the front. **Res_type** records represent each individual residential and sub-residential account, with each account given a unique identifier. Each city is assigned a specific **Municipal_id**. Records indicate water consumption for a specific account during a single billing period. Consumption is related to both Account and Location by referencing an **Account_No** to an **Address**.

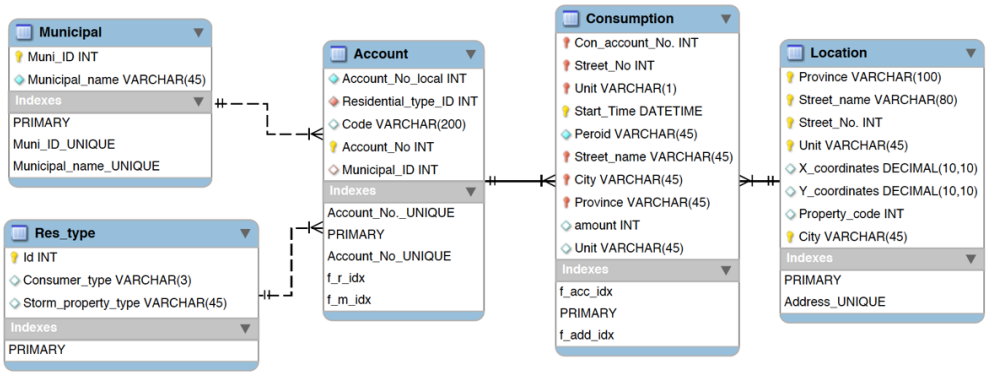



**Figure 4. 1 Water demand database EER**

The structure of the database is designed in an account-transaction style to facilitate expansion. As more records being imported, computations will need to be distributed to across more machines given the increased load on the algorithm and framework.

## 4.2 Requirements and Specifications

The highest-level requirement is to be able to uniquely define the relationship between residential water consumption with water price and weather for the City of Waterloo. To the degree that records are short in duration, measurements of water consumption are obtained only to the nearest cubic meter, weather is approximated from temperature and precipitation measurements, and other ambient process such as policy are ignored, this requirement can never be met. However, we can reasonably define the relationship during the time interval from 2007-2016 during which data is available. This requirement has been met by replicating the analysis done using Excel as recorded by Enouy (2018). Therefore, a more accurate requirement should be stated as: validation by virtue of implementing the algorithm from Enouy (2018). This validation effort is acceptable given the following assumptions.

1) The Excel spreadsheet implementation is consistent with the Enouy (2018).
2) Enouy (2018) correctly defines the mathematical methodology.

### 4.2.1 Prototyping

One important step of working on specifications with clients and trying to test the algorithm at the same time is to build a prototype in shortest time.

Matlab is a perfect option for building a prototype since it has a complete a library of plotting diagrams, a variety of options for a least square problem solver. One of the important steps of this algorithm is to generate probability density function parameters by solving a least square problem. The quality of the generated results significantly depends selecting a correct algorithm for the solver and properly applying it. Because the non-linear least square algorithm applied by excel is exclusive, it might be a better choice for us to make full use of the optimization library in Matlab, instead of implementing a random non-linear least square algorithm in another language.

In addition, by enabling the client to monitor the whole process of converging to the desired result from starting point through each iteration, it gives them a better vision on how to quantify efficiency requirements on the program.

```
                                        Norm of      First-order
 Iteration  Func-count     f(x)          step          optimality
     0          6         22.7491                         38.5
     1         12         1.46905         5.6481          12.7
     2         18       0.0246356        2.47207           0.9
     3         24     0.000396985       0.615079        0.0683
     4         30     0.000134524       0.253638         0.015
     5         36     0.000101864       0.230588       0.00295
     6         42     0.000100205      0.0684746      0.000156
     7         48       0.0001002     0.00341715      2.89e-07

Local minimum found.

Optimization completed because the size of the gradient is less than
the selected value of the optimality tolerance.

<stopping criteria details>


Using polynomial series Extension by trust-region-reflective
PDF: y = (exp(-(0.797192 + 0.290841x + 1/2*tan(53.124658/180*pi)x^2 + -0.192689x^3+ 0.018420x^4)))/18.421904
MSE_ob: 0.000006
MSE_m: 0.000241
```

**Figure 4. 2 Results generated by Matlab**

The plotting library helps visualize the results since visualizing the quality of fitting results is used as part of the tests.

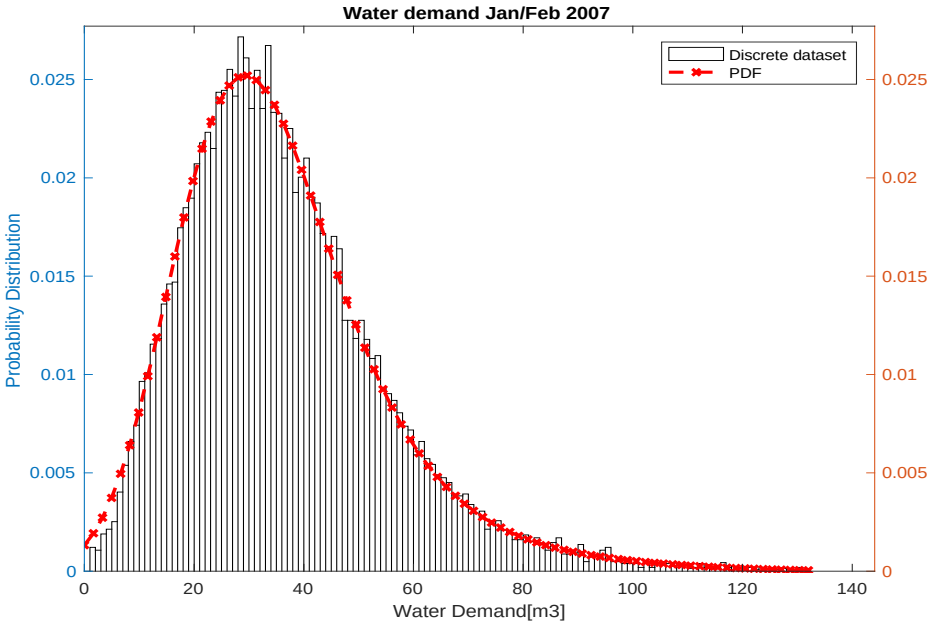



**Figure 4. 3 Histogram fitting plotted by Matlab**

The user interface of the prototype is constructed in the form of question-answer data entry. User can only enter the options available to them on the screen. The advantage of this type of interface is that it restricts the options of user input, making the program more robust, and most importantly, easier and faster to program.

The expansion of the requirement might lead to a much larger and complicated program, to the extent where Matlab code might fail to meet the minimum efficiency level that is required. Therefore, the prototype is most likely to be discarded once a complete specification is formed.

### 4.2.2 Client Writing Out Black Box Test Cases

The spreadsheet implementation is done by the client not only to test the algorithm, but also to be used as a guide for a software engineer. But from a software engineer's perspective, it is better to be used as black box test cases. The published paper should be the only material to a software engineer can refer to, which motivates the software engineer to go through every detail in the paper. Any confusion should be solved by communicating directly with the client instead of referring to the excel sheet.

Value of many parameters can be chosen alternatively from the excel implementation such as, step, starting point for a least square. Those might lead to different efficiency and accuracy. Taking different paths might be helpful for finding the best combination of both efficiency and accuracy.

As mentioned in the previous section, the implementation in excel is broken down to several steps. Each step should be set up as a single test case throughout the whole implementation. Failure to pass the test does not directly indicate the implementation of the prototype is wrong. Instead, it proves

1) prototype is wrong
2) or implementation of the excel is inconsistent with paper
3) or the least square algorithm in excel (Non-linear GRG) behaves differently from those in Matlab (trust-region and levernberg-marquardt)

Number 3 should only be a matter of choice if they both generate ideal results.

Through this approach, errors resulting from inconsistencies can be very well eliminated.

### 4.2.3 Communication through Sequence Diagram

The implementations in excel are done under very limited conditions. In these case, even wrong implementations can occasionally achieve ideal results but fail to do so when being applied in more cases. Since the algorithm has yet to be proven a robust one, errors can possibly result from both the limitation of the algorithm or mistakes in the paper, when excel implementation is proven to be consistent with the paper.

It is impossible for software engineer to capture ideas in a client's mind that are inconsistent with what was written in the paper. But good communication is still crucial when the client is trying to be convinced that the algorithm is properly implemented.

Asking a client to go through thousands of lines of code is tedious and unreasonable.

Multiple scenarios of sequence diagrams can describe the process in more understandable ways, which give the client a complete picture on whether the program is implemented in the way they desire and finally rule out the possibility that the implementation is wrong.

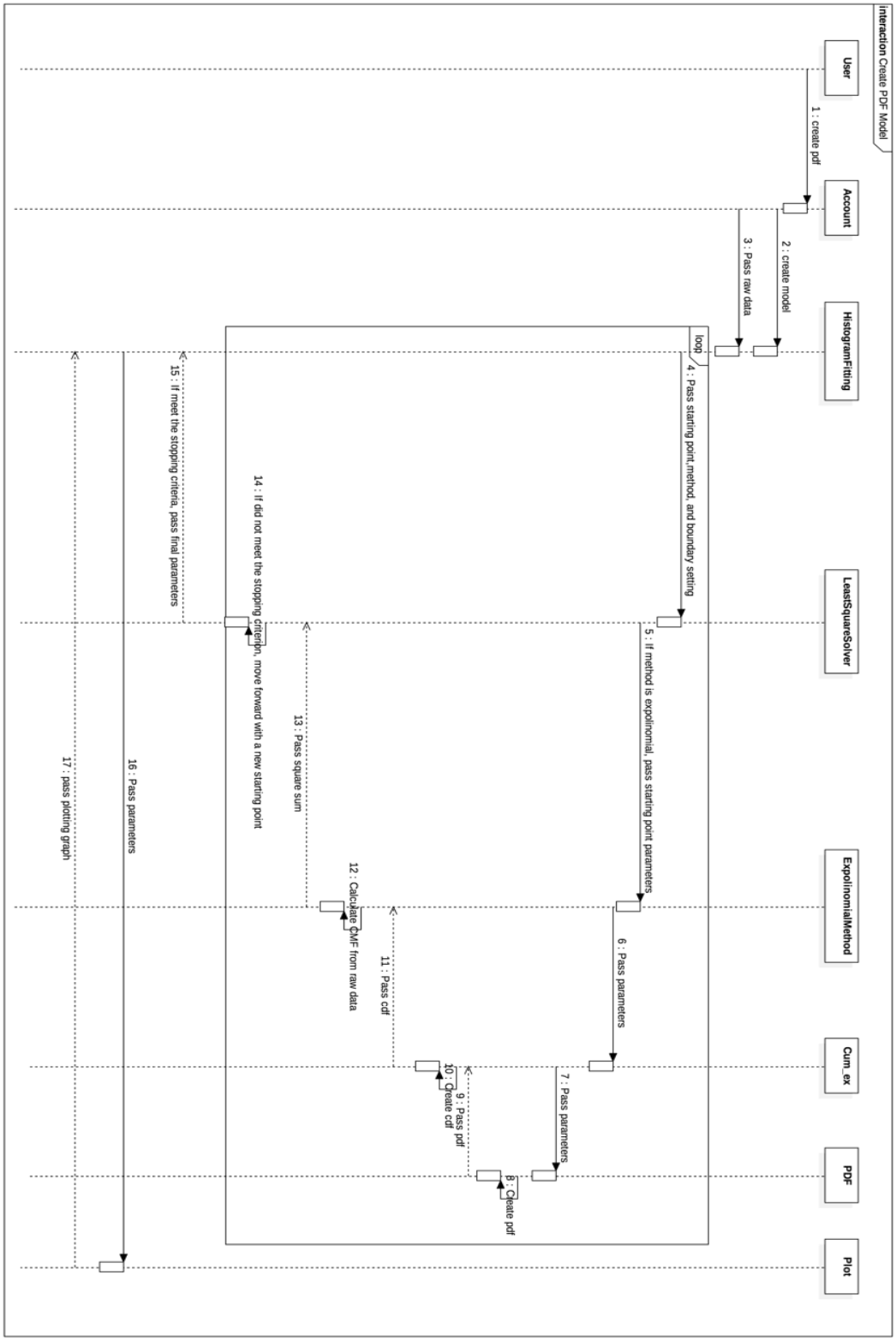



**Figure 4. 4 Sequence Diagram for creating PDF**

# 5 CONCLUSION

This work is developed based on Dr. Enouy's theory, applying the methodology into water utilities, learning the how residential and commercial water demand evolves as water price and weather changes. The analysis focuses on the application of this methodology into residential and commercial water consumption in the City of Waterloo, exploring its potential value as a software in terms of efficiency and robustness.

The results of histogram fitting for residential water consumption are mostly compromised by many outliers at the tail of the histogram. Not only those high volumes of outliers fall far out of the range that a single family can consume, they also had a huge impact on the efficiency and fitting results. The data culling method is proven to be helpful to solve these issues. It is also refined in this work for two reasons.

1) As an anomaly detection technique, able to distinguish a heavy-tailed distribution from one having many outliers.
2) To find a proper culling ratio that minimizes the loss of information and maximize the performance improvements.

This refined analysis proves that using 4 as a culling ratio according to Dr.Enouy (2018) is appropriate. It also classified commercial water consumption in each billing period as a heavy-tailed distribution, instead of culling data that aren't outliers, resulting in significant loss of information.

This work also demonstrates how Trust-Region is advantageous over Levernberg-Marquardt, regarding both performance and results' quality, and how control function parameter scaling can slightly improve efficiency in the case of fitting residential water consumption data, which could be a huge improvement when fitting some cities with much shorter billing periods. Adjustments are made for fitting commercial water demand, by fitting data in the lognormal of median-relative space. More intellectual schemes for shape recognition and data transformation should be made, when dealing distributions abnormal forms.

The regression analysis applies curvilinear regression, aiming to find out high-order relationships between dependent and independent variables. It is shown that control function

parameters tend to depend on one variable, either weather score or real water price, while mean statistics, median and standard deviation are dependent on both.

The last chapter talks about employing multiple strategies from the software industry, with the goal to guarantee a more reliable, efficient transfer from academic work to real-life application.

# REFERENCES

American Water Works Association (AWWA). (2012). Buried No Longer: Confronting America's Water Infrastructure Challenge. http://www.awwa.org/legislation-regulation/issues/infrastructure-financing.aspx. Accessed 2/26/2015.

Berghen, F. V. (2004). Levenberg-Marquardt algorithms vs trust region algorithms. *IRIDIA, Université Libre de Bruxelles*.

Coleman, Thomas F., and Yuying Li. "An interior trust region approach for nonlinear minimization subject to bounds." SIAM Journal on optimization 6.2 (1996): 418-445.

Gavin, H. (2016). The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems. 2011. *Available Online from: http://people. duke. edu/~ hpgavin/ce281/lm. pdf..(Accessed 15 Sept 2015)*.

Raftelis Financial Consultants, Inc., 2012 Water and Wastewater Rate Survey - Book

Robert Enouy (2018). An Investigation into Water Consumption Data Using Parametric Probability Density Functions. UWSpace.

# APPENDIX

Table 1: Maximum of water consumption in each billing period vs Median

| Billing Periods | Most consumption(m3) | Median |
|---|---|---|
| **2007-01-01 0:00** | 498 | 33 |
| **2007-03-01 0:00** | 734 | 31 |
| **2007-05-01 0:00** | 585 | 33 |
| **2007-07-01 0:00** | 339 | 40 |
| **2007-09-01 0:00** | 471 | 36 |
| **2007-11-01 0:00** | 737 | 33 |
| **2008-01-01 0:00** | 2560 | 33 |
| **2008-03-01 0:00** | 2080 | 33 |
| **2008-05-01 0:00** | 1850 | 32 |
| **2008-07-01 0:00** | 9983 | 35 |
| **2008-09-01 0:00** | 2540 | 32 |
| **2008-11-01 0:00** | 2450 | 31 |
| **2009-01-01 0:00** | 1600 | 31 |
| **2009-03-01 0:00** | 1600 | 31 |
| **2009-05-01 0:00** | 1680 | 31 |
| **2009-07-01 0:00** | 2060 | 33 |
| **2009-09-01 0:00** | 3110 | 31 |
| **2009-11-01 0:00** | 3030 | 31 |
| **2010-01-01 0:00** | 3066 | 31 |
| **2010-03-01 0:00** | 20006 | 31 |
| **2010-05-01 0:00** | 10011 | 30 |
| **2010-07-01 0:00** | 2026 | 33 |
| **2010-09-01 0:00** | 3060 | 30 |
| **2010-11-01 0:00** | 2830 | 30 |
| **2011-01-01 0:00** | 1661 | 30 |
| **2011-03-01 0:00** | 2090 | 29 |
| **2011-05-01 0:00** | 2746 | 30 |
| **2011-07-01 0:00** | 2220 | 32 |
| **2011-09-01 0:00** | 2770 | 31 |
| **2011-11-01 0:00** | 4320 | 29 |
| **2012-01-01 0:00** | 3536 | 27 |
| **2012-03-01 0:00** | 3626 | 28 |
| **2012-05-01 0:00** | 3470 | 28 |
| **2012-07-01 0:00** | 3054 | 33 |

| | | |
|---|---|---|
| **2012-09-01 0:00** | 3629 | 31 |
| **2012-11-01 0:00** | 5367 | 28 |
| **2013-01-01 0:00** | 4155 | 27 |
| **2013-03-01 0:00** | 3309 | 28 |
| **2013-05-01 0:00** | 3736 | 27 |
| **2013-07-01 0:00** | 4006 | 29 |
| **2013-09-01 0:00** | 3563 | 28 |
| **2013-11-01 0:00** | 4786 | 28 |
| **2014-01-01 0:00** | 5255 | 28 |
| **2014-03-01 0:00** | 4008 | 27 |
| **2014-05-01 0:00** | 12021 | 27 |
| **2014-07-01 0:00** | 4397 | 28 |
| **2014-09-01 0:00** | 11167 | 27 |
| **2014-11-01 0:00** | 6992 | 27 |
| **2015-01-01 0:00** | 12410 | 26 |
| **2015-03-01 0:00** | 8245 | 27 |
| **2015-05-01 0:00** | 8550 | 27 |
| **2015-07-01 0:00** | 10607 | 28 |
| **2015-09-01 0:00** | 8859 | 27 |
| **2015-11-01 0:00** | 10205 | 27 |
| **2016-01-01 0:00** | 7550 | 25 |
| **2016-03-01 0:00** | 8428 | 25 |
| **2016-05-01 0:00** | 9877 | 26 |
| **2016-07-01 0:00** | 10661 | 30 |
| **2016-09-01 0:00** | 11107 | 28 |
| **2016-11-01 0:00** | 10240 | 26 |

Table 2: Median under different culling ratios

| Billing Periods | Actual Median | Median when n = 4 | Median when n = 5 |
|---|---|---|---|
| **2007-01-01 0:00** | 33 | 33 | 33 |
| **2007-03-01 0:00** | 31 | 31 | 31 |
| **2007-05-01 0:00** | 33 | 32 | 32 |
| **2007-07-01 0:00** | 40 | 40 | 40 |
| **2007-09-01 0:00** | 36 | 36 | 36 |
| **2007-11-01 0:00** | 33 | 33 | 33 |
| **2008-01-01 0:00** | 33 | 33 | 33 |
| **2008-03-01 0:00** | 33 | 32 | 33 |
| **2008-05-01 0:00** | 32 | 32 | 32 |
| **2008-07-01 0:00** | 35 | 34 | 35 |
| **2008-09-01 0:00** | 32 | 32 | 32 |
| **2008-11-01 0:00** | 31 | 31 | 31 |
| **2009-01-01 0:00** | 31 | 31 | 31 |
| **2009-03-01 0:00** | 31 | 31 | 31 |
| **2009-05-01 0:00** | 31 | 31 | 31 |
| **2009-07-01 0:00** | 33 | 33 | 33 |
| **2009-09-01 0:00** | 31 | 31 | 31 |
| **2009-11-01 0:00** | 31 | 31 | 31 |
| **2010-01-01 0:00** | 31 | 31 | 31 |
| **2010-03-01 0:00** | 31 | 31 | 31 |
| **2010-05-01 0:00** | 30 | 30 | 30 |
| **2010-07-01 0:00** | 33 | 33 | 33 |
| **2010-09-01 0:00** | 30 | 30 | 30 |
| **2010-11-01 0:00** | 30 | 30 | 30 |
| **2011-01-01 0:00** | 30 | 30 | 30 |
| **2011-03-01 0:00** | 29 | 29 | 29 |
| **2011-05-01 0:00** | 30 | 30 | 30 |
| **2011-07-01 0:00** | 32 | 32 | 32 |
| **2011-09-01 0:00** | 31 | 31 | 31 |
| **2011-11-01 0:00** | 29 | 29 | 29 |
| **2012-01-01 0:00** | 27 | 27 | 27 |
| **2012-03-01 0:00** | 28 | 28 | 28 |
| **2012-05-01 0:00** | 28 | 28 | 28 |
| **2012-07-01 0:00** | 33 | 33 | 33 |
| **2012-09-01 0:00** | 31 | 31 | 31 |

| | | | |
|---|---|---|---|
| **2012-11-01 0:00** | 28 | 28 | 28 |
| **2013-01-01 0:00** | 27 | 27 | 27 |
| **2013-03-01 0:00** | 28 | 28 | 28 |
| **2013-05-01 0:00** | 27 | 27 | 27 |
| **2013-07-01 0:00** | 29 | 29 | 29 |
| **2013-09-01 0:00** | 28 | 28 | 28 |
| **2013-11-01 0:00** | 28 | 28 | 28 |
| **2014-01-01 0:00** | 28 | 28 | 28 |
| **2014-03-01 0:00** | 27 | 27 | 27 |
| **2014-05-01 0:00** | 27 | 27 | 27 |
| **2014-07-01 0:00** | 28 | 28 | 28 |
| **2014-09-01 0:00** | 27 | 27 | 27 |
| **2014-11-01 0:00** | 27 | 27 | 27 |
| **2015-01-01 0:00** | 26 | 25 | 26 |
| **2015-03-01 0:00** | 27 | 26 | 26 |
| **2015-05-01 0:00** | 27 | 26 | 27 |
| **2015-07-01 0:00** | 28 | 28 | 28 |
| **2015-09-01 0:00** | 27 | 27 | 27 |
| **2015-11-01 0:00** | 27 | 26 | 26 |
| **2016-01-01 0:00** | 25 | 25 | 25 |
| **2016-03-01 0:00** | 25 | 25 | 25 |
| **2016-05-01 0:00** | 26 | 26 | 26 |
| **2016-07-01 0:00** | 30 | 30 | 30 |
| **2016-09-01 0:00** | 28 | 27 | 27 |
| **2016-11-01 0:00** | 26 | 26 | 26 |

Table 3: Amount of accounts left under different culling ratios

| Billing Periods | Number of Accounts/n=4 | Number of Accounts/n=5 | Number of Accounts/n=6 | Number of Accounts/n=7 | Number of Accounts/n=10 |
|---|---|---|---|---|---|
| **2007-01-01 0:00** | 22643 | 20616 | 20653 | 20672 | 20681 |
| **2007-03-01 0:00** | 22643 | 20677 | 20744 | 20768 | 20793 |
| **2007-05-01 0:00** | 22643 | 20606 | 20657 | 20676 | 20688 |
| **2007-07-01 0:00** | 22643 | 21344 | 21438 | 21468 | 21487 |
| **2007-09-01 0:00** | 22643 | 21817 | 21944 | 21976 | 22015 |
| **2007-11-01 0:00** | 22643 | 21492 | 21569 | 21600 | 21626 |
| **2008-01-01 0:00** | 24908 | 22470 | 22528 | 22552 | 22576 |
| **2008-03-01 0:00** | 24908 | 22509 | 22626 | 22668 | 22710 |
| **2008-05-01 0:00** | 24908 | 21851 | 21899 | 21926 | 21947 |
| **2008-07-01 0:00** | 24908 | 22847 | 22964 | 23002 | 23028 |
| **2008-09-01 0:00** | 24908 | 23971 | 24089 | 24131 | 24191 |
| **2008-11-01 0:00** | 24908 | 23413 | 23500 | 23528 | 23556 |
| **2009-01-01 0:00** | 27223 | 24663 | 24737 | 24770 | 24789 |
| **2009-03-01 0:00** | 27223 | 24693 | 24800 | 24841 | 24876 |
| **2009-05-01 0:00** | 27223 | 25179 | 25294 | 25334 | 25364 |
| **2009-07-01 0:00** | 27223 | 24265 | 24365 | 24405 | 24433 |
| **2009-09-01 0:00** | 27223 | 26351 | 26518 | 26576 | 26638 |
| **2009-11-01 0:00** | 27223 | 26803 | 26901 | 26935 | 26969 |
| **2010-01-01 0:00** | 27457 | 26828 | 26925 | 26966 | 26999 |
| **2010-03-01 0:00** | 27457 | 26649 | 26786 | 26846 | 26900 |
| **2010-05-01 0:00** | 27457 | 25173 | 25253 | 25274 | 25303 |
| **2010-07-01 0:00** | 27457 | 26531 | 26712 | 26778 | 26846 |
| **2010-09-01 0:00** | 27457 | 25131 | 25293 | 25353 | 25418 |
| **2010-11-01 0:00** | 27457 | 26876 | 27025 | 27080 | 27141 |
| **2011-01-01 0:00** | 27538 | 23657 | 23720 | 23741 | 23776 |
| **2011-03-01 0:00** | 27538 | 26828 | 26945 | 26986 | 27037 |
| **2011-05-01 0:00** | 27538 | 26711 | 26915 | 26999 | 27080 |
| **2011-07-01 0:00** | 27538 | 26708 | 26861 | 26909 | 26970 |

| | | | | | |
|---|---|---|---|---|---|
| **2011-09-01 0:00** | 27538 | 23454 | 23603 | 23659 | 23727 |
| **2011-11-01 0:00** | 27538 | 26847 | 26992 | 27041 | 27107 |
| **2012-01-01 0:00** | 27622 | 20986 | 21053 | 21085 | 21109 |
| **2012-03-01 0:00** | 27622 | 26873 | 26971 | 27014 | 27038 |
| **2012-05-01 0:00** | 27622 | 26948 | 27040 | 27065 | 27098 |
| **2012-07-01 0:00** | 27622 | 26653 | 26840 | 26919 | 26984 |
| **2012-09-01 0:00** | 27622 | 26697 | 26922 | 27039 | 27135 |
| **2012-11-01 0:00** | 27622 | 27013 | 27125 | 27166 | 27219 |
| **2013-01-01 0:00** | 27744 | 27088 | 27179 | 27217 | 27262 |
| **2013-03-01 0:00** | 27744 | 26959 | 27038 | 27063 | 27105 |
| **2013-05-01 0:00** | 27744 | 26988 | 27070 | 27105 | 27151 |
| **2013-07-01 0:00** | 27744 | 26889 | 27027 | 27073 | 27121 |
| **2013-09-01 0:00** | 27744 | 26926 | 27109 | 27188 | 27273 |
| **2013-11-01 0:00** | 27744 | 27223 | 27317 | 27367 | 27412 |
| **2014-01-01 0:00** | 27907 | 27264 | 27348 | 27373 | 27412 |
| **2014-03-01 0:00** | 27907 | 27048 | 27139 | 27173 | 27216 |
| **2014-05-01 0:00** | 27907 | 26983 | 27078 | 27107 | 27152 |
| **2014-07-01 0:00** | 27907 | 26903 | 27056 | 27123 | 27188 |
| **2014-09-01 0:00** | 27907 | 27074 | 27214 | 27278 | 27343 |
| **2014-11-01 0:00** | 27907 | 27268 | 27369 | 27409 | 27460 |
| **2015-01-01 0:00** | 28042 | 27282 | 27401 | 27429 | 27479 |
| **2015-03-01 0:00** | 28042 | 27123 | 27217 | 27252 | 27329 |
| **2015-05-01 0:00** | 28042 | 27154 | 27269 | 27303 | 27356 |
| **2015-07-01 0:00** | 28042 | 27096 | 27234 | 27279 | 27347 |
| **2015-09-01 0:00** | 28042 | 27141 | 27310 | 27391 | 27468 |
| **2015-11-01 0:00** | 28042 | 27301 | 27430 | 27483 | 27543 |
| **2016-01-01 0:00** | 26306 | 25839 | 25924 | 25961 | 26010 |
| **2016-03-01 0:00** | 26306 | 25746 | 25832 | 25866 | 25902 |
| **2016-05-01 0:00** | 26306 | 25874 | 25951 | 25986 | 26029 |
| **2016-07-01 0:00** | 26306 | 25728 | 25874 | 25942 | 26007 |
| **2016-09-01 0:00** | 26306 | 25681 | 25889 | 25971 | 26063 |
| **2016-11-01 0:00** | 26306 | 25837 | 25952 | 25996 | 26049 |

Table 4: Percentage of accounts left under different culling ratios

| Billing Periods | Percentage of Accounts Left/n=4 | Percentage of Accounts Left/n=5 | Percentage of Accounts Left/n=6 | Percentage of Accounts Left/n=10 |
|---|---|---|---|---|
| **2007-01-01 0:00** | 99.68% | 99.85% | 99.95% | 99.99% |
| **2007-03-01 0:00** | 99.41% | 99.73% | 99.85% | 99.97% |
| **2007-05-01 0:00** | 99.59% | 99.84% | 99.93% | 99.99% |
| **2007-07-01 0:00** | 99.33% | 99.77% | 99.91% | 100.00% |
| **2007-09-01 0:00** | 99.09% | 99.67% | 99.81% | 99.99% |
| **2007-11-01 0:00** | 99.34% | 99.69% | 99.84% | 99.96% |
| **2008-01-01 0:00** | 99.49% | 99.74% | 99.85% | 99.96% |
| **2008-03-01 0:00** | 99.08% | 99.59% | 99.78% | 99.96% |
| **2008-05-01 0:00** | 99.55% | 99.77% | 99.89% | 99.99% |
| **2008-07-01 0:00** | 99.18% | 99.69% | 99.86% | 99.97% |
| **2008-09-01 0:00** | 99.05% | 99.53% | 99.71% | 99.95% |
| **2008-11-01 0:00** | 99.38% | 99.75% | 99.86% | 99.98% |
| **2009-01-01 0:00** | 99.47% | 99.77% | 99.90% | 99.98% |
| **2009-03-01 0:00** | 99.20% | 99.63% | 99.80% | 99.94% |
| **2009-05-01 0:00** | 99.24% | 99.69% | 99.85% | 99.96% |
| **2009-07-01 0:00** | 99.28% | 99.69% | 99.86% | 99.97% |
| **2009-09-01 0:00** | 98.88% | 99.51% | 99.73% | 99.96% |
| **2009-11-01 0:00** | 99.32% | 99.69% | 99.81% | 99.94% |
| **2010-01-01 0:00** | 99.29% | 99.65% | 99.80% | 99.93% |
| **2010-03-01 0:00** | 99.01% | 99.52% | 99.74% | 99.94% |
| **2010-05-01 0:00** | 99.43% | 99.74% | 99.83% | 99.94% |
| **2010-07-01 0:00** | 98.76% | 99.43% | 99.68% | 99.93% |
| **2010-09-01 0:00** | 98.78% | 99.42% | 99.65% | 99.91% |
| **2010-11-01 0:00** | 98.94% | 99.49% | 99.69% | 99.92% |
| **2011-01-01 0:00** | 99.42% | 99.68% | 99.77% | 99.92% |
| **2011-03-01 0:00** | 99.16% | 99.59% | 99.74% | 99.93% |
| **2011-05-01 0:00** | 98.57% | 99.32% | 99.63% | 99.93% |
| **2011-07-01 0:00** | 98.97% | 99.53% | 99.71% | 99.94% |
| **2011-09-01 0:00** | 98.76% | 99.39% | 99.63% | 99.91% |
| **2011-11-01 0:00** | 98.92% | 99.46% | 99.64% | 99.88% |
| **2012-01-01 0:00** | 99.30% | 99.61% | 99.76% | 99.88% |
| **2012-03-01 0:00** | 99.26% | 99.62% | 99.78% | 99.87% |
| **2012-05-01 0:00** | 99.34% | 99.68% | 99.77% | 99.89% |

| | | | | |
|---|---|---|---|---|
| **2012-07-01 0:00** | 98.68% | 99.37% | 99.67% | 99.91% |
| **2012-09-01 0:00** | 98.23% | 99.06% | 99.49% | 99.84% |
| **2012-11-01 0:00** | 99.08% | 99.49% | 99.64% | 99.84% |
| **2013-01-01 0:00** | 99.20% | 99.53% | 99.67% | 99.84% |
| **2013-03-01 0:00** | 99.31% | 99.61% | 99.70% | 99.85% |
| **2013-05-01 0:00** | 99.26% | 99.57% | 99.69% | 99.86% |
| **2013-07-01 0:00** | 98.96% | 99.47% | 99.64% | 99.82% |
| **2013-09-01 0:00** | 98.53% | 99.20% | 99.49% | 99.80% |
| **2013-11-01 0:00** | 99.15% | 99.49% | 99.68% | 99.84% |
| **2014-01-01 0:00** | 99.25% | 99.56% | 99.65% | 99.79% |
| **2014-03-01 0:00** | 99.16% | 99.49% | 99.62% | 99.78% |
| **2014-05-01 0:00** | 99.17% | 99.51% | 99.62% | 99.79% |
| **2014-07-01 0:00** | 98.78% | 99.35% | 99.59% | 99.83% |
| **2014-09-01 0:00** | 98.81% | 99.32% | 99.55% | 99.79% |
| **2014-11-01 0:00** | 99.10% | 99.47% | 99.61% | 99.80% |
| **2015-01-01 0:00** | 99.08% | 99.52% | 99.62% | 99.80% |
| **2015-03-01 0:00** | 98.98% | 99.32% | 99.45% | 99.73% |
| **2015-05-01 0:00** | 99.05% | 99.47% | 99.60% | 99.79% |
| **2015-07-01 0:00** | 98.88% | 99.38% | 99.55% | 99.80% |
| **2015-09-01 0:00** | 98.56% | 99.18% | 99.47% | 99.75% |
| **2015-11-01 0:00** | 98.88% | 99.35% | 99.54% | 99.76% |
| **2016-01-01 0:00** | 99.10% | 99.43% | 99.57% | 99.76% |
| **2016-03-01 0:00** | 99.15% | 99.48% | 99.61% | 99.75% |
| **2016-05-01 0:00** | 99.20% | 99.49% | 99.63% | 99.79% |
| **2016-07-01 0:00** | 98.72% | 99.28% | 99.54% | 99.79% |
| **2016-09-01 0:00** | 98.31% | 99.10% | 99.42% | 99.77% |
| **2016-11-01 0:00** | 98.94% | 99.38% | 99.55% | 99.75% |

Table 5: Performance and results, original vs culled

| Billing Periods | Number of Iterations/original | Residual/original | Number of Iterations/n=4 | Residual/n=4 |
|---|---|---|---|---|
| **2007-01-01 0:00** | 9 | 1.33E-05 | 7 | 8.84E-06 |
| **2007-03-01 0:00** | 7 | 2.25E-06 | 6 | 7.62E-06 |
| **2007-05-01 0:00** | 9 | 6.33E-06 | 9 | 1.08E-05 |
| **2007-07-01 0:00** | 8 | 3.97E-06 | 8 | 3.97E-06 |
| **2007-09-01 0:00** | 8 | 1.12E-05 | 9 | 1.38E-05 |
| **2007-11-01 0:00** | 9 | 3.16E-06 | 7 | 4.63E-06 |
| **2008-01-01 0:00** | 15 | 1.31E-05 | 9 | 6.45E-06 |
| **2008-03-01 0:00** | 17 | 1.60E-05 | 8 | 5.78E-06 |
| **2008-05-01 0:00** | 9 | 1.81E-06 | 8 | 1.67E-06 |
| **2008-07-01 0:00** | 31 | 4.98E-05 | 8 | 2.19E-06 |
| **2008-09-01 0:00** | 22 | 1.78E-05 | 9 | 5.71E-06 |
| **2008-11-01 0:00** | 10 | 4.09E-06 | 8 | 4.02E-06 |
| **2009-01-01 0:00** | 10 | 1.96E-06 | 7 | 5.93E-06 |
| **2009-03-01 0:00** | 12 | 8.77E-06 | 8 | 8.41E-06 |
| **2009-05-01 0:00** | 10 | 9.55E-06 | 7 | 6.80E-06 |
| **2009-07-01 0:00** | 10 | 6.00E-06 | 7 | 3.94E-06 |
| **2009-09-01 0:00** | 28 | 2.27E-06 | 9 | 7.27E-06 |
| **2009-11-01 0:00** | 15 | 1.26E-06 | 8 | 3.02E-06 |
| **2010-01-01 0:00** | 25 | 1.90E-05 | 8 | 2.97E-06 |
| **2010-03-01 0:00** | 40 | 1.36E-06 | 8 | 4.64E-06 |
| **2010-05-01 0:00** | 46 | 4.49E-05 | 8 | 6.28E-06 |
| **2010-07-01 0:00** | 9 | 5.56E-06 | 9 | 8.38E-06 |
| **2010-09-01 0:00** | 30 | 3.83E-06 | 9 | 4.66E-06 |
| **2010-11-01 0:00** | 17 | 2.76E-06 | 9 | 3.15E-06 |
| **2011-01-01 0:00** | 17 | 6.46E-07 | 7 | 4.38E-06 |
| **2011-03-01 0:00** | 16 | 6.10E-06 | 8 | 1.62E-06 |
| **2011-05-01 0:00** | 32 | 1.49E-05 | 13 | 4.65E-06 |
| **2011-07-01 0:00** | 30 | 1.59E-05 | 8 | 3.64E-06 |
| **2011-09-01 0:00** | 27 | 3.07E-05 | 10 | 6.68E-06 |
| **2011-11-01 0:00** | 24 | 2.80E-06 | 10 | 5.73E-06 |

| | | | | |
|---|---|---|---|---|
| **2012-01-01 0:00** | 29 | 2.21E-06 | 7 | 5.08E-06 |
| **2012-03-01 0:00** | 28 | 1.29E-06 | 7 | 3.87E-06 |
| **2012-05-01 0:00** | 29 | 1.09E-05 | 7 | 8.04E-06 |
| **2012-07-01 0:00** | 27 | 1.58E-05 | 8 | 8.02E-06 |
| **2012-09-01 0:00** | 27 | 3.28E-05 | 13 | 1.32E-05 |
| **2012-11-01 0:00** | 43 | 2.70E-06 | 9 | 1.27E-05 |
| **2013-01-01 0:00** | 26 | 2.87E-05 | 7 | 3.86E-06 |
| **2013-03-01 0:00** | 26 | 7.07E-07 | 8 | 2.07E-06 |
| **2013-05-01 0:00** | 29 | 2.25E-05 | 8 | 5.43E-06 |
| **2013-07-01 0:00** | 25 | 3.78E-05 | 8 | 9.02E-06 |
| **2013-09-01 0:00** | 30 | 4.97E-05 | 10 | 4.02E-06 |
| **2013-11-01 0:00** | 28 | 1.09E-05 | 7 | 5.75E-06 |
| **2014-01-01 0:00** | 35 | 1.51E-05 | 8 | 5.90E-06 |
| **2014-03-01 0:00** | 51 | 2.59E-05 | 7 | 4.80E-06 |
| **2014-05-01 0:00** | 63 | 1.25E-04 | 7 | 3.60E-06 |
| **2014-07-01 0:00** | 29 | 2.40E-06 | 9 | 9.35E-06 |
| **2014-09-01 0:00** | 46 | 7.36E-05 | 9 | 4.84E-06 |
| **2014-11-01 0:00** | 51 | 3.85E-06 | 8 | 4.06E-06 |
| **2015-01-01 0:00** | 22 | 8.97E-05 | 7 | 3.41E-06 |
| **2015-03-01 0:00** | 53 | 2.20E-05 | 8 | 2.19E-06 |
| **2015-05-01 0:00** | 43 | 7.59E-06 | 8 | 2.22E-06 |
| **2015-07-01 0:00** | 76 | 2.04E-04 | 9 | 6.40E-06 |
| **2015-09-01 0:00** | 58 | 1.23E-04 | 10 | 1.05E-05 |
| **2015-11-01 0:00** | 26 | 8.81E-07 | 8 | 2.66E-06 |
| **2016-01-01 0:00** | 67 | 3.46E-06 | 9 | 4.31E-06 |
| **2016-03-01 0:00** | 63 | 2.92E-06 | 8 | 6.81E-06 |
| **2016-05-01 0:00** | 22 | 3.78E-06 | 7 | 1.87E-06 |
| **2016-07-01 0:00** | 51 | 2.75E-04 | 8 | 1.53E-05 |
| **2016-09-01 0:00** | 37 | 2.69E-04 | 11 | 6.80E-06 |
| **2016-11-01 0:00** | 17 | 2.92E-06 | 8 | 6.89E-06 |

Table 5: Performance and results, Unscaled control function parameters vs Scaled

| Billing Periods | Number of Iterations/Starts at [0,0,45,0,0] | Residual/Starts at [0,0,45,0,0] | Number of Iterations/Starts at [0,0,1,0,0] | Residual/Starts [0,0,1,0,0] |
|---|---|---|---|---|
| **2007-01-01 0:00** | 7 | 8.84E-06 | 8 | 5.58E-06 |
| **2007-03-01 0:00** | 6 | 7.62E-06 | 7 | 1.56E-06 |
| **2007-05-01 0:00** | 9 | 1.08E-05 | 8 | 6.25E-06 |
| **2007-07-01 0:00** | 8 | 3.97E-06 | 8 | 3.76E-06 |
| **2007-09-01 0:00** | 9 | 1.38E-05 | 8 | 8.35E-06 |
| **2007-11-01 0:00** | 7 | 4.63E-06 | 7 | 3.38E-06 |
| **2008-01-01 0:00** | 9 | 6.45E-06 | 7 | 2.65E-06 |
| **2008-03-01 0:00** | 8 | 5.78E-06 | 8 | 1.17E-05 |
| **2008-05-01 0:00** | 8 | 1.67E-06 | 7 | 4.83E-06 |
| **2008-07-01 0:00** | 8 | 2.19E-06 | 7 | 6.77E-06 |
| **2008-09-01 0:00** | 9 | 5.71E-06 | 7 | 7.49E-06 |
| **2008-11-01 0:00** | 8 | 4.02E-06 | 7 | 3.63E-06 |
| **2009-01-01 0:00** | 7 | 5.93E-06 | 7 | 4.51E-06 |
| **2009-03-01 0:00** | 8 | 8.41E-06 | 8 | 7.14E-06 |
| **2009-05-01 0:00** | 7 | 6.80E-06 | 7 | 2.98E-06 |
| **2009-07-01 0:00** | 7 | 3.94E-06 | 7 | 7.67E-06 |
| **2009-09-01 0:00** | 9 | 7.27E-06 | 7 | 4.28E-06 |
| **2009-11-01 0:00** | 8 | 3.02E-06 | 6 | 3.00E-06 |
| **2010-01-01 0:00** | 8 | 2.97E-06 | 7 | 3.54E-06 |
| **2010-03-01 0:00** | 8 | 4.64E-06 | 7 | 3.46E-06 |
| **2010-05-01 0:00** | 8 | 6.28E-06 | 7 | 3.01E-06 |
| **2010-07-01 0:00** | 9 | 8.38E-06 | 7 | 6.16E-06 |
| **2010-09-01 0:00** | 9 | 4.66E-06 | 8 | 4.53E-06 |
| **2010-11-01 0:00** | 9 | 3.15E-06 | 8 | 2.72E-06 |
| **2011-01-01 0:00** | 7 | 4.38E-06 | 7 | 3.78E-06 |
| **2011-03-01 0:00** | 8 | 1.62E-06 | 7 | 2.04E-06 |
| **2011-05-01 0:00** | 13 | 4.65E-06 | 8 | 2.80E-06 |
| **2011-07-01 0:00** | 8 | 3.64E-06 | 7 | 5.68E-06 |
| **2011-09-01 0:00** | 10 | 6.68E-06 | 8 | 9.24E-06 |
| **2011-11-01 0:00** | 10 | 5.73E-06 | 8 | 5.57E-06 |
| **2012-01-01 0:00** | 7 | 5.08E-06 | 6 | 4.00E-06 |
| **2012-03-01 0:00** | 7 | 3.87E-06 | 7 | 3.14E-06 |
| **2012-05-01 0:00** | 7 | 8.04E-06 | 7 | 1.78E-06 |
| **2012-07-01 0:00** | 8 | 8.02E-06 | 7 | 7.33E-06 |

| | | | | |
|---|---|---|---|---|
| **2012-09-01 0:00** | 13 | 1.32E-05 | 7 | 3.50E-06 |
| **2012-11-01 0:00** | 9 | 1.27E-05 | 8 | 6.21E-06 |
| **2013-01-01 0:00** | 7 | 3.86E-06 | 7 | 6.37E-06 |
| **2013-03-01 0:00** | 8 | 2.07E-06 | 7 | 1.73E-06 |
| **2013-05-01 0:00** | 8 | 5.43E-06 | 7 | 4.44E-06 |
| **2013-07-01 0:00** | 8 | 9.02E-06 | 7 | 6.72E-06 |
| **2013-09-01 0:00** | 10 | 4.02E-06 | 7 | 9.21E-06 |
| **2013-11-01 0:00** | 7 | 5.75E-06 | 8 | 6.81E-06 |
| **2014-01-01 0:00** | 8 | 5.90E-06 | 8 | 3.58E-06 |
| **2014-03-01 0:00** | 7 | 4.80E-06 | 6 | 3.91E-06 |
| **2014-05-01 0:00** | 7 | 3.60E-06 | 7 | 5.97E-06 |
| **2014-07-01 0:00** | 9 | 9.35E-06 | 7 | 3.23E-06 |
| **2014-09-01 0:00** | 9 | 4.84E-06 | 7 | 4.86E-06 |
| **2014-11-01 0:00** | 8 | 4.06E-06 | 8 | 5.32E-06 |
| **2015-01-01 0:00** | 7 | 3.41E-06 | 8 | 3.82E-06 |
| **2015-03-01 0:00** | 8 | 2.19E-06 | 7 | 2.99E-06 |
| **2015-05-01 0:00** | 8 | 2.22E-06 | 7 | 4.01E-06 |
| **2015-07-01 0:00** | 9 | 6.40E-06 | 7 | 2.41E-06 |
| **2015-09-01 0:00** | 10 | 1.05E-05 | 7 | 1.23E-06 |
| **2015-11-01 0:00** | 8 | 2.66E-06 | 7 | 5.41E-06 |
| **2016-01-01 0:00** | 9 | 4.31E-06 | 8 | 1.66E-06 |
| **2016-03-01 0:00** | 8 | 6.81E-06 | 7 | 3.38E-06 |
| **2016-05-01 0:00** | 7 | 1.87E-06 | 8 | 6.75E-06 |
| **2016-07-01 0:00** | 8 | 1.53E-05 | 7 | 3.32E-06 |
| **2016-09-01 0:00** | 11 | 6.80E-06 | 7 | 3.23E-06 |
| **2016-11-01 0:00** | 8 | 6.89E-06 | 8 | 6.57E-06 |
| **Average** | 8.28 | 5.84E-06 | 7.27 | 4.68E-06 |

Table 6: Trust-Region vs Levenberg–Marquardt

| Billing Periods | Trust-Region iterations | Trust-Region residuals | Levenberg–Marquardt algorithm iterations | Levenberg–Marquardt algorithm residuals |
|---|---|---|---|---|
| 2007-01-01 0:00 | 8 | 5.58E-06 | 43 | 5.58E-06 |
| 2007-03-01 0:00 | 7 | 1.56E-06 | 37 | 1.56E-06 |
| 2007-05-01 0:00 | 8 | 6.25E-06 | 43 | 6.25E-06 |
| 2007-07-01 0:00 | 8 | 3.76E-06 | 43 | 3.76E-06 |
| 2007-09-01 0:00 | 8 | 8.35E-06 | 43 | 8.35E-06 |
| 2007-11-01 0:00 | 7 | 3.38E-06 | 37 | 3.38E-06 |
| 2008-01-01 0:00 | 7 | 2.65E-06 | 37 | 2.65E-06 |
| 2008-03-01 0:00 | 8 | 1.17E-05 | 43 | 1.17E-05 |
| 2008-05-01 0:00 | 7 | 4.83E-06 | 37 | 4.83E-06 |
| 2008-07-01 0:00 | 7 | 6.77E-06 | 37 | 6.77E-06 |
| 2008-09-01 0:00 | 7 | 7.49E-06 | 37 | 7.49E-06 |
| 2008-11-01 0:00 | 7 | 3.63E-06 | 37 | 3.63E-06 |
| 2009-01-01 0:00 | 7 | 4.51E-06 | 37 | 4.51E-06 |
| 2009-03-01 0:00 | 8 | 7.14E-06 | 43 | 7.14E-06 |
| 2009-05-01 0:00 | 7 | 2.98E-06 | 37 | 2.98E-06 |
| 2009-07-01 0:00 | 7 | 7.67E-06 | 37 | 7.67E-06 |
| 2009-09-01 0:00 | 7 | 4.28E-06 | 37 | 4.28E-06 |
| 2009-11-01 0:00 | 6 | 3.00E-06 | 37 | 3.00E-06 |
| 2010-01-01 0:00 | 7 | 3.54E-06 | 37 | 3.54E-06 |
| 2010-03-01 0:00 | 7 | 3.46E-06 | 37 | 3.46E-06 |
| 2010-05-01 0:00 | 7 | 3.01E-06 | 37 | 3.01E-06 |
| 2010-07-01 0:00 | 7 | 6.16E-06 | 37 | 6.16E-06 |
| 2010-09-01 0:00 | 8 | 4.53E-06 | 43 | 4.53E-06 |
| 2010-11-01 0:00 | 8 | 2.72E-06 | 43 | 2.72E-06 |
| 2011-01-01 0:00 | 7 | 3.78E-06 | 37 | 3.78E-06 |
| 2011-03-01 0:00 | 7 | 2.04E-06 | 37 | 2.04E-06 |
| 2011-05-01 0:00 | 8 | 2.80E-06 | 43 | 2.80E-06 |
| 2011-07-01 0:00 | 7 | 5.68E-06 | 37 | 5.68E-06 |
| 2011-09-01 0:00 | 8 | 9.24E-06 | 43 | 9.24E-06 |
| 2011-11-01 0:00 | 8 | 5.57E-06 | 43 | 5.57E-06 |
| 2012-01-01 0:00 | 6 | 4.00E-06 | 37 | 4.00E-06 |
| 2012-03-01 0:00 | 7 | 3.14E-06 | 37 | 3.14E-06 |
| 2012-05-01 0:00 | 7 | 1.78E-06 | 37 | 1.78E-06 |
| 2012-07-01 0:00 | 7 | 7.33E-06 | 37 | 7.33E-06 |
| 2012-09-01 0:00 | 7 | 3.50E-06 | 37 | 3.50E-06 |
| 2012-11-01 0:00 | 8 | 6.21E-06 | 43 | 6.21E-06 |

| | | | | |
|---|---|---|---|---|
| 2013-01-01 0:00 | 7 | 6.37E-06 | 37 | 6.37E-06 |
| 2013-03-01 0:00 | 7 | 1.73E-06 | 37 | 1.73E-06 |
| 2013-05-01 0:00 | 7 | 4.44E-06 | 37 | 4.44E-06 |
| 2013-07-01 0:00 | 7 | 6.72E-06 | 37 | 6.72E-06 |
| 2013-09-01 0:00 | 7 | 9.21E-06 | 37 | 9.21E-06 |
| 2013-11-01 0:00 | 8 | 6.81E-06 | 43 | 6.81E-06 |
| 2014-01-01 0:00 | 8 | 3.58E-06 | 43 | 3.58E-06 |
| 2014-03-01 0:00 | 6 | 3.91E-06 | 37 | 3.91E-06 |
| 2014-05-01 0:00 | 7 | 5.97E-06 | 37 | 5.97E-06 |
| 2014-07-01 0:00 | 7 | 3.23E-06 | 37 | 3.23E-06 |
| 2014-09-01 0:00 | 7 | 4.86E-06 | 37 | 4.86E-06 |
| 2014-11-01 0:00 | 8 | 5.32E-06 | 43 | 5.32E-06 |
| 2015-01-01 0:00 | 8 | 3.82E-06 | 43 | 3.82E-06 |
| 2015-03-01 0:00 | 7 | 2.99E-06 | 37 | 2.99E-06 |
| 2015-05-01 0:00 | 7 | 4.01E-06 | 37 | 4.01E-06 |
| 2015-07-01 0:00 | 7 | 2.41E-06 | 37 | 2.41E-06 |
| 2015-09-01 0:00 | 7 | 1.23E-06 | 37 | 1.23E-06 |
| 2015-11-01 0:00 | 7 | 5.41E-06 | 37 | 5.41E-06 |
| 2016-01-01 0:00 | 8 | 1.66E-06 | 43 | 1.66E-06 |
| 2016-03-01 0:00 | 7 | 3.38E-06 | 37 | 3.38E-06 |
| 2016-05-01 0:00 | 8 | 6.75E-06 | 43 | 6.75E-06 |
| 2016-07-01 0:00 | 7 | 3.32E-06 | 37 | 3.32E-06 |
| 2016-09-01 0:00 | 7 | 3.23E-06 | 37 | 3.23E-06 |
| 2016-11-01 0:00 | 8 | 6.57E-06 | 43 | 6.57E-06 |
| Average | 7.27 | 4.68E-06 | 38.9 | 4.68E-06 |

Table 7: Trust-Region at different starting position

| Billing Periods | Trust-Region iterations/start at[0,0,1,0,0] | Trust-Region iterations/start at[10,10,2,10,10] | Trust-Region residuals/start at[0,0,1,0,0] | Trust-Region residuals/start at[10,10,2,10,10] |
|---|---|---|---|---|
| 2007-01-01 0:00 | 8 | 126 | 5.58E-06 | 6.60E-06 |
| 2007-03-01 0:00 | 7 | 102 | 1.56E-06 | 1.62E-06 |
| 2007-05-01 0:00 | 8 | 125 | 6.25E-06 | 6.69E-06 |
| 2007-07-01 0:00 | 8 | 157 | 3.76E-06 | 5.60E-06 |
| 2007-09-01 0:00 | 8 | 108 | 8.35E-06 | 1.19E-05 |
| 2007-11-01 0:00 | 7 | 119 | 3.38E-06 | 4.14E-06 |
| 2008-01-01 0:00 | 7 | 124 | 2.65E-06 | 3.21E-06 |
| 2008-03-01 0:00 | 8 | 119 | 1.17E-05 | 1.24E-05 |
| 2008-05-01 0:00 | 7 | 102 | 4.83E-06 | 5.11E-06 |
| 2008-07-01 0:00 | 7 | 92 | 6.77E-06 | 7.67E-06 |
| 2008-09-01 0:00 | 7 | 97 | 7.49E-06 | 9.21E-06 |
| 2008-11-01 0:00 | 7 | 87 | 3.63E-06 | 4.12E-06 |
| 2009-01-01 0:00 | 7 | 99 | 4.51E-06 | 5.95E-06 |
| 2009-03-01 0:00 | 8 | 89 | 7.14E-06 | 8.76E-06 |
| 2009-05-01 0:00 | 7 | 94 | 2.98E-06 | 4.34E-06 |
| 2009-07-01 0:00 | 7 | 121 | 7.67E-06 | 8.48E-06 |
| 2009-09-01 0:00 | 7 | 101 | 4.28E-06 | 5.45E-06 |
| 2009-11-01 0:00 | 6 | 98 | 3.00E-06 | 3.30E-06 |
| 2010-01-01 0:00 | 7 | 89 | 3.54E-06 | 3.84E-06 |
| 2010-03-01 0:00 | 7 | 95 | 3.46E-06 | 3.28E-06 |
| 2010-05-01 0:00 | 7 | 100 | 3.01E-06 | 2.99E-06 |
| 2010-07-01 0:00 | 7 | 90 | 6.16E-06 | 5.21E-06 |
| 2010-09-01 0:00 | 8 | 99 | 4.53E-06 | 4.12E-06 |
| 2010-11-01 0:00 | 8 | 92 | 2.72E-06 | 3.34E-06 |
| 2011-01-01 0:00 | 7 | 110 | 3.78E-06 | 4.82E-06 |
| 2011-03-01 0:00 | 7 | 161 | 2.04E-06 | 2.58E-06 |
| 2011-05-01 0:00 | 8 | 89 | 2.80E-06 | 3.03E-06 |
| 2011-07-01 0:00 | 7 | 89 | 5.68E-06 | 8.73E-06 |
| 2011-09-01 0:00 | 8 | 96 | 9.24E-06 | 1.13E-05 |
| 2011-11-01 0:00 | 8 | 158 | 5.57E-06 | 5.65E-06 |
| 2012-01-01 0:00 | 6 | 199 | 4.00E-06 | 4.54E-06 |
| 2012-03-01 0:00 | 7 | 171 | 3.14E-06 | 4.05E-06 |
| 2012-05-01 0:00 | 7 | 179 | 1.78E-06 | 2.41E-06 |
| 2012-07-01 0:00 | 7 | 91 | 7.33E-06 | 1.10E-05 |
| 2012-09-01 0:00 | 7 | 94 | 3.50E-06 | 6.06E-06 |

| | | | | |
|---|---|---|---|---|
| 2012-11-01 0:00 | 8 | 179 | 6.21E-06 | 7.68E-06 |
| 2013-01-01 0:00 | 7 | 193 | 6.37E-06 | 7.61E-06 |
| 2013-03-01 0:00 | 7 | 170 | 1.73E-06 | 2.49E-06 |
| 2013-05-01 0:00 | 7 | 195 | 4.44E-06 | 5.84E-06 |
| 2013-07-01 0:00 | 7 | 166 | 6.72E-06 | 6.69E-06 |
| 2013-09-01 0:00 | 7 | 181 | 9.21E-06 | 1.17E-05 |
| 2013-11-01 0:00 | 8 | 192 | 6.81E-06 | 8.60E-06 |
| 2014-01-01 0:00 | 8 | 173 | 3.58E-06 | 2.94E-06 |
| 2014-03-01 0:00 | 6 | 194 | 3.91E-06 | 4.64E-06 |
| 2014-05-01 0:00 | 7 | 184 | 5.97E-06 | 7.55E-06 |
| 2014-07-01 0:00 | 7 | 182 | 3.23E-06 | 3.12E-06 |
| 2014-09-01 0:00 | 7 | 184 | 4.86E-06 | 4.61E-06 |
| 2014-11-01 0:00 | 8 | 196 | 5.32E-06 | 7.22E-06 |
| 2015-01-01 0:00 | 8 | 195 | 3.82E-06 | 3.27E-06 |
| 2015-03-01 0:00 | 7 | 182 | 2.99E-06 | 2.54E-06 |
| 2015-05-01 0:00 | 7 | 181 | 4.01E-06 | 3.61E-06 |
| 2015-07-01 0:00 | 7 | 178 | 2.41E-06 | 3.69E-06 |
| 2015-09-01 0:00 | 7 | 177 | 1.23E-06 | 1.96E-06 |
| 2015-11-01 0:00 | 7 | 182 | 5.41E-06 | 4.73E-06 |
| 2016-01-01 0:00 | 8 | 197 | 1.66E-06 | 1.58E-06 |
| 2016-03-01 0:00 | 7 | 247 | 3.38E-06 | 4.49E-06 |
| 2016-05-01 0:00 | 8 | 195 | 6.75E-06 | 8.08E-06 |
| 2016-07-01 0:00 | 7 | 91 | 3.32E-06 | 5.24E-06 |
| 2016-09-01 0:00 | 7 | 184 | 3.23E-06 | 5.25E-06 |
| 2016-11-01 0:00 | 8 | 184 | 6.57E-06 | 7.90E-06 |
| Average | 7.27 | 142.4 | 4.68E-06 | 5.58E-06 |

Table 8: Levenberg-Marquardt at different starting position

| Billing Periods | Levenberg-Marquardt iterations/start at[0,0,1,0,0] | Levenberg-Marquardt iterations/start at[10,10,2,10,10] | Levenberg-Marquardt residuals/start at[0,0,1,0,0] | Levenberg-Marquardt residuals/start at[10,10,2,10,10] |
|---|---|---|---|---|
| 2007-01-01 0:00 | 43 | 780 | 5.58E-06 | 6.60E-06 |
| 2007-03-01 0:00 | 37 | 634 | 1.56E-06 | 1.62E-06 |
| 2007-05-01 0:00 | 43 | 685 | 6.25E-06 | 6.69E-06 |
| 2007-07-01 0:00 | 43 | 563 | 3.76E-06 | 5.60E-06 |
| 2007-09-01 0:00 | 43 | 676 | 8.35E-06 | 1.19E-05 |
| 2007-11-01 0:00 | 37 | 733 | 3.38E-06 | 4.14E-06 |
| 2008-01-01 0:00 | 37 | 769 | 2.65E-06 | 3.21E-06 |
| 2008-03-01 0:00 | 43 | 663 | 1.17E-05 | 1.24E-05 |
| 2008-05-01 0:00 | 37 | 641 | 4.83E-06 | 5.11E-06 |
| 2008-07-01 0:00 | 37 | 695 | 6.77E-06 | 7.67E-06 |
| 2008-09-01 0:00 | 37 | 565 | 7.49E-06 | 9.21E-06 |
| 2008-11-01 0:00 | 37 | 570 | 3.63E-06 | 4.12E-06 |
| 2009-01-01 0:00 | 37 | 708 | 4.51E-06 | 5.95E-06 |
| 2009-03-01 0:00 | 43 | 640 | 7.14E-06 | 8.76E-06 |
| 2009-05-01 0:00 | 37 | 641 | 2.98E-06 | 4.34E-06 |
| 2009-07-01 0:00 | 37 | 678 | 7.67E-06 | 8.48E-06 |
| 2009-09-01 0:00 | 37 | 618 | 4.28E-06 | 5.45E-06 |
| 2009-11-01 0:00 | 37 | 589 | 3.00E-06 | 3.30E-06 |
| 2010-01-01 0:00 | 37 | 610 | 3.54E-06 | 3.84E-06 |
| 2010-03-01 0:00 | 37 | 580 | 3.46E-06 | 3.28E-06 |
| 2010-05-01 0:00 | 37 | 666 | 3.01E-06 | 2.99E-06 |
| 2010-07-01 0:00 | 37 | 599 | 6.16E-06 | 5.21E-06 |
| 2010-09-01 0:00 | 43 | 608 | 4.53E-06 | 4.12E-06 |
| 2010-11-01 0:00 | 43 | 655 | 2.72E-06 | 3.34E-06 |
| 2011-01-01 0:00 | 37 | 664 | 3.78E-06 | 4.82E-06 |
| 2011-03-01 0:00 | 37 | 690 | 2.04E-06 | 2.58E-06 |
| 2011-05-01 0:00 | 43 | 611 | 2.80E-06 | 3.03E-06 |
| 2011-07-01 0:00 | 37 | 573 | 5.68E-06 | 8.73E-06 |
| 2011-09-01 0:00 | 43 | 633 | 9.24E-06 | 1.13E-05 |
| 2011-11-01 0:00 | 43 | 680 | 5.57E-06 | 5.65E-06 |
| 2012-01-01 0:00 | 37 | 714 | 4.00E-06 | 4.54E-06 |
| 2012-03-01 0:00 | 37 | 711 | 3.14E-06 | 4.05E-06 |
| 2012-05-01 0:00 | 37 | 711 | 1.78E-06 | 2.41E-06 |
| 2012-07-01 0:00 | 37 | 605 | 7.33E-06 | 1.10E-05 |

| | | | | |
|---|---|---|---|---|
| 2012-09-01 0:00 | 37 | 602 | 3.50E-06 | 6.06E-06 |
| 2012-11-01 0:00 | 43 | 713 | 6.21E-06 | 7.68E-06 |
| 2013-01-01 0:00 | 37 | 703 | 6.37E-06 | 7.61E-06 |
| 2013-03-01 0:00 | 37 | 696 | 1.73E-06 | 2.49E-06 |
| 2013-05-01 0:00 | 37 | 707 | 4.44E-06 | 5.84E-06 |
| 2013-07-01 0:00 | 37 | 708 | 6.72E-06 | 6.69E-06 |
| 2013-09-01 0:00 | 37 | 677 | 9.21E-06 | 1.17E-05 |
| 2013-11-01 0:00 | 43 | 774 | 6.81E-06 | 8.60E-06 |
| 2014-01-01 0:00 | 43 | 723 | 3.58E-06 | 2.94E-06 |
| 2014-03-01 0:00 | 37 | 743 | 3.91E-06 | 4.64E-06 |
| 2014-05-01 0:00 | 37 | 711 | 5.97E-06 | 7.55E-06 |
| 2014-07-01 0:00 | 37 | 716 | 3.23E-06 | 3.12E-06 |
| 2014-09-01 0:00 | 37 | 710 | 4.86E-06 | 4.61E-06 |
| 2014-11-01 0:00 | 43 | 777 | 5.32E-06 | 7.22E-06 |
| 2015-01-01 0:00 | 43 | 723 | 3.82E-06 | 3.27E-06 |
| 2015-03-01 0:00 | 37 | 718 | 2.99E-06 | 2.54E-06 |
| 2015-05-01 0:00 | 37 | 728 | 4.01E-06 | 3.61E-06 |
| 2015-07-01 0:00 | 37 | 683 | 2.41E-06 | 3.69E-06 |
| 2015-09-01 0:00 | 37 | 739 | 1.23E-06 | 1.96E-06 |
| 2015-11-01 0:00 | 37 | 686 | 5.41E-06 | 4.73E-06 |
| 2016-01-01 0:00 | 43 | 723 | 1.66E-06 | 1.58E-06 |
| 2016-03-01 0:00 | 37 | 871 | 3.38E-06 | 4.49E-06 |
| 2016-05-01 0:00 | 43 | 722 | 6.75E-06 | 8.08E-06 |
| 2016-07-01 0:00 | 37 | 651 | 3.32E-06 | 5.24E-06 |
| 2016-09-01 0:00 | 37 | 750 | 3.23E-06 | 5.25E-06 |
| 2016-11-01 0:00 | 43 | 808 | 6.57E-06 | 7.90E-06 |
| Average | 7.27 | 682 | 4.68E-06 | 5.58E-06 |

Table 8: Performance and result, LogY vs Y

| Billing Periods | LogY Space/Iterations | LogY Space/Residuals | Y Space/Iterations | Y Space/Residuals |
|---|---|---|---|---|
| **2007-01-01 0:00** | 12 | 2.86E-05 | 9 | 4.24E-05 |
| **2007-03-01 0:00** | 10 | 2.76E-05 | 10 | 3.96E-05 |
| **2007-05-01 0:00** | 14 | 2.46E-05 | 8 | 1.12E-04 |
| **2007-07-01 0:00** | 13 | 8.38E-05 | 10 | 3.34E-05 |
| **2007-09-01 0:00** | 11 | 1.65E-05 | 9 | 5.22E-05 |
| **2007-11-01 0:00** | 12 | 1.08E-05 | 8 | 2.68E-05 |
| **2008-01-01 0:00** | 9 | 2.29E-05 | 77 | 1.87E-03 |
| **2008-03-01 0:00** | 9 | 3.86E-05 | 57 | 2.71E-03 |
| **2008-05-01 0:00** | 8 | 1.79E-05 | 56 | 7.29E-03 |
| **2008-07-01 0:00** | 8 | 1.36E-05 | 34 | 5.35E-03 |
| **2008-09-01 0:00** | 10 | 3.88E-05 | 46 | 6.18E-03 |
| **2008-11-01 0:00** | 10 | 2.60E-05 | 29 | 6.35E-03 |
| **2009-01-01 0:00** | 8 | 2.27E-05 | 44 | 5.66E-03 |
| **2009-03-01 0:00** | 8 | 2.50E-05 | 21 | 2.15E-02 |
| **2009-05-01 0:00** | 8 | 1.64E-05 | 46 | 5.37E-03 |
| **2009-07-01 0:00** | 8 | 1.41E-05 | 51 | 6.28E-03 |
| **2009-09-01 0:00** | 8 | 1.44E-05 | 54 | 6.57E-03 |
| **2009-11-01 0:00** | 9 | 1.14E-05 | 48 | 5.40E-03 |
| **2010-01-01 0:00** | 9 | 2.47E-05 | 50 | 4.88E-03 |
| **2010-03-01 0:00** | 9 | 1.56E-05 | 51 | 5.35E-03 |
| **2010-05-01 0:00** | 8 | 1.98E-05 | 50 | 5.10E-03 |
| **2010-07-01 0:00** | 9 | 1.94E-05 | 43 | 4.57E-03 |
| **2010-09-01 0:00** | 9 | 2.84E-05 | 38 | 5.29E-03 |
| **2010-11-01 0:00** | 8 | 2.31E-05 | 47 | 5.54E-03 |
| **2011-01-01 0:00** | 9 | 4.02E-05 | 33 | 6.34E-03 |
| **2011-03-01 0:00** | 9 | 2.57E-05 | 43 | 5.48E-03 |
| **2011-05-01 0:00** | 8 | 1.77E-05 | 51 | 5.11E-03 |
| **2011-07-01 0:00** | 9 | 3.49E-05 | 47 | 5.01E-03 |
| **2011-09-01 0:00** | 9 | 3.65E-05 | 52 | 4.51E-03 |
| **2011-11-01 0:00** | 9 | 2.67E-05 | 55 | 5.26E-03 |
| **2012-01-01 0:00** | 8 | 5.50E-05 | 39 | 5.02E-03 |
| **2012-03-01 0:00** | 8 | 2.15E-05 | 52 | 5.36E-03 |
| **2012-05-01 0:00** | 8 | 1.86E-05 | 39 | 4.82E-03 |
| **2012-07-01 0:00** | 9 | 2.59E-05 | 47 | 4.59E-03 |
| **2012-09-01 0:00** | 8 | 1.19E-05 | 46 | 4.89E-03 |

| | | | | |
|---|---|---|---|---|
| **2012-11-01 0:00** | 8 | 2.03E-05 | 44 | 5.09E-03 |
| **2013-01-01 0:00** | 7 | 2.03E-05 | 46 | 4.69E-03 |
| **2013-03-01 0:00** | 9 | 2.36E-05 | 40 | 4.62E-03 |
| **2013-05-01 0:00** | 8 | 2.85E-05 | 46 | 4.67E-03 |
| **2013-07-01 0:00** | 9 | 1.88E-05 | 46 | 4.27E-03 |
| **2013-09-01 0:00** | 8 | 2.41E-05 | 48 | 4.15E-03 |
| **2013-11-01 0:00** | 10 | 3.05E-05 | 36 | 4.99E-03 |
| **2014-01-01 0:00** | 9 | 3.14E-05 | 45 | 4.46E-03 |
| **2014-03-01 0:00** | 9 | 2.56E-05 | 49 | 4.82E-03 |
| **2014-05-01 0:00** | 8 | 3.29E-05 | 48 | 4.38E-03 |
| **2014-07-01 0:00** | 8 | 3.30E-05 | 39 | 4.12E-03 |
| **2014-09-01 0:00** | 8 | 3.33E-05 | 39 | 5.31E-03 |
| **2014-11-01 0:00** | 9 | 2.92E-05 | 46 | 4.30E-03 |
| **2015-01-01 0:00** | 9 | 2.64E-05 | 42 | 4.22E-03 |
| **2015-03-01 0:00** | 9 | 2.30E-05 | 45 | 4.41E-03 |
| **2015-05-01 0:00** | 8 | 1.72E-05 | 49 | 4.87E-03 |
| **2015-07-01 0:00** | 8 | 3.57E-05 | 42 | 5.24E-03 |
| **2015-09-01 0:00** | 8 | 2.36E-05 | 41 | 4.12E-03 |
| **2015-11-01 0:00** | 8 | 2.47E-05 | 46 | 4.51E-03 |
| **2016-01-01 0:00** | 9 | 2.67E-05 | 44 | 4.33E-03 |
| **2016-03-01 0:00** | 9 | 2.93E-05 | 48 | 4.26E-03 |
| **2016-05-01 0:00** | 8 | 3.52E-05 | 44 | 4.54E-03 |
| **2016-07-01 0:00** | 8 | 1.45E-05 | 42 | 4.75E-03 |
| **2016-09-01 0:00** | 8 | 3.58E-05 | 56 | 4.70E-03 |
| **2016-11-01 0:00** | 9 | 5.26E-05 | 43 | 4.15E-03 |
| Average | 8.87 | 2.6593E-05 | 41.9 | 0.0047 |