Usage Analysis for Smart Meter Management

Hongfei Li
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA, 10598
liho@us.ibm.com

Dongping Fang
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA, 10598
dfang@us.ibm.com

Shilpa Mahatma
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA, 10598
mahatma@us.ibm.com

Arun Hampapur
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA, 10598
arunh@us.ibm.com

Abstract — **Smart meters gather utility usage data, such as water, electricity and gas readings, by remote reporting. The flood of usage data obtained from each meter in real-time or near real-time enable data analytics and optimization tool to support smart meter management. Predictive usage analytics can provide significant benefits to both the utilities and customers. We propose statistical approaches for meter anomaly detection, usage demand forecasting and association analysis for utility companies. These analyses provide efficient ways to detect malfunctioning meters, optimize water supply in the future and understand the association factors that drive meter failures and water demand. We illustrate our methodology using the automated meter reading (AMR) database from a water utility customer.**

Keywords – Anomaly Detection, Association Analysis, Automated Meter Reading (AMR), Demand Forecasting, Smart Meter Management, Usage Analysis

I. INTRODUCTION

A. Smart Meter Management

Smart meters gather utility usage data, such as water, electricity and gas readings, by remote reporting and bring a flood of data to utility companies and a source of intelligence. The need to manage the data, and subsequently transform it into actionable business intelligence, creates challenges for utilities implementing smart metering. The usage data obtained from each meter in real-time or near real-time enable data analytics and optimization tool to support smart meter management.

Optimal use of analytics for smart meter management can help utilities improve customer relationships, through more regular and targeted demand response programs, boosting customer loyalty and minimizing wasted marketing spend; achieve greater network reliability and resilience, with real-time, automated updates about equipment status and operations. Faults and outages can be isolated and addressed more quickly and effectively. This improved responsiveness, in turn, helps to build enhanced and durable customer relationships [1]. Specifically, predictive usage analytics can provide significant benefits to both the utilities and customers from three perspectives: meter anomaly detection, usage demand forecasting and association analysis. Anomaly detection tool identifies unusual patterns of energy or water consumption using historical meter readings and then to detect malfunctioning meters or potential theft behaviors of customers. Demand forecasting tool shapes demand patterns in the future and allows operators to be more aware of their energy usage which provides the basis of energy supply optimization. Association analysis is integrated in anomaly detection and demand forecasting tools by incorporating external factors, such as customer attributes and weather information, to explain the root cause of anomalies and demand prediction.

Figure 1 describes the overall picture of predictive-usage-analysis-based smart meter management. Frequent reading collection is prerequisite for usage analytics. Additional information, such as customer properties, asset condition data and expertise knowledge, enrich the data sources by supplying more comprehensive information for decision making. The data sources serve as the inputs of the predictive usage analytics toolkit which integrates different functionalities to support smart meter management. Analytics can be applied to project future energy demand and revenue and to help plan capital upgrades of the infrastructure. In terms of operations, detection of malfunctioning meters help with field testing and optimization of work management. Demand forecasting supports energy supply planning and scheduling.

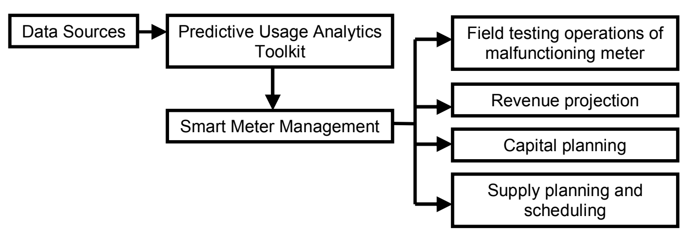

Figure 1: Usage-analysis-based smart meter management

B. Automated Meter Reading (AMR)

At the center of understanding water usage is the ability to measure, analyze, model, and predict the consumption of water at a fairly granular timescale. One of the key challenges for this exercise is the availability of data. Traditionally, meter reading has been a manual process in which the meter reader had to physically look at the dial. Our work is based on an ongoing engagement with a water utility customer [2, 3]. In 2002, the water utility company initiated an effort to replace traditional water meters with automated meters. It is believed that they were one of the first water utilities to implement a fixed network Automated Meter Reading (AMR) system in the

978-1-4577-1591-4/11/$26.00 ©2011 IEEE

United States. The system uses both radio and cell phone technology to upload data at the meter to the database twice daily. This provided more accurate data and eliminated estimated billing. Since the data no longer needed to be read manually, manpower, fuel and vehicle maintenance costs were drastically reduced.

The water utility company developed the award-wining application as a free service to proactively notify customers of high water use - including unknown household leaks, sprinklers accidentally left running or ruptured washing machine hoses. This application has generated more than 18,000 notifications to customers since January 2006 and it improves the customer service. Beyond the current use of AMR system, we proposed a much wider range of applications of usage analysis on smart meter readings to support smart meter management.

This paper describes the application of water usage predictive analytics in smart meter management to deliver business value. The implementation of these analytics using software products has been integrated to an end-to-end software solution for analytics-driven asset management (ADAM) [4].

This paper will be organized as follows. Section II gives an overview of predictive usage analytics toolkit followed by data description. The last part of the Section II outlines the statistical modeling implemented in predictive usage analytics. Section III illustrates the application of the toolkit using the use case of the water utility customer. The paper ends with discussion in Section IV.

II. PREDICTIVE USAGE ANALYTICS TOOLKIT

A. Overview

We developed predictive analytical toolkit for smart meter management based on the AMR database where the daily water usage were recorded and external information such as customer attributes and weather conditions are also available. SME provides domain expert knowledge which supplements unavailable or unaccounted data information. These data sources provide inputs to our modeling toolkit.

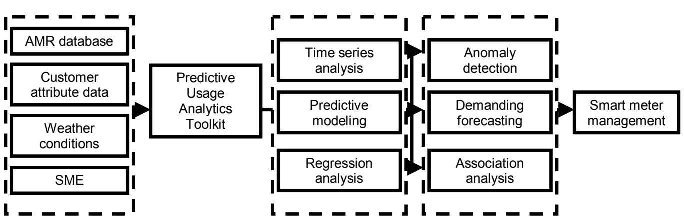

Figure 2: Overview of predictive analytics toolkit

The customer and usage analytics utilizes statistical approaches including time series analysis, predictive modeling and regression analysis. It allows the utility to better understand the usage of water, its customer base, and its revenue and then forecast future water usage. Specifically, this application provides the following three functionalities:

1) Anomaly Detection

Usage anomalies can occur as a result of intentional customer fraud, meter malfunctioning, leakage or reading errors. In any case, usage anomaly impacts the bottom line of the utility. Anomaly detection can assume many forms, including detection of demographic inconsistency, which is the case when a customer's usage does not fit the typical usage pattern of the customer demographic, or when there is usage inconsistency in which the current usage does not fit the historical usage pattern for the customer. Several types of meter malfunctioning have been identified using meter readings, such as meter silting, bypass and usage sudden changes. For example, sand sediment in meters causes water readings record less than actually consumed. Gradually decreasing trend in water consumption over time provides an indicator of meter silting problem. Water theft and by-pass problem can be identified by a sudden drop or level shift in usages.

The time series model in Section C is developed to detect each type of the anomalies in one model which provides a guideline for utility companies for either field testing or billing adjustment.

2) Demand Forecasting

This is a critical function for utilities. Forecasting usage takes into account of seasonal usage patterns, peak usage, long term usage trends and project usage at an individual and customer segment level. This capability becomes the basis for water supply optimization, revenue forecasting, infrastructure planning and early warnings into emerging problems. We applied time series multivariate regression models and predicted water usage in the future. This model incorporates the temporal correlation, seasonal patterns and external impacts, such as weather.

3) Association Analysis

Anomaly detection and demand forecasting are crucial to support the operations of smart meter management. However, it is also interesting to operators to understand what factors drive the anomalies and demand. This tool provides different functions in different applications. In meter anomaly detection, customer attributes, such as customer types, residential or commercial, and geographical information, and meter attributes, such as meter age and size, will be associated with malfunctioning meters to explain the correlation between the external factors and the meter failures. In demand forecasting, in addition to customer and meter attributes, weather information and seasonal impact also play an important role in usage, thus it is critical to explore their association.

B. Data Sources

In the application of the water utility customer, the AMR data were recorded twice a day, morning and evening, at individual customer level from January 1st, 2008 to June 30th, 2011. We take the evening readings as the customer's daily reading. Let Z_t denote the reading of day t. Since the reading data are accumulative over days, i.e., $Z_t \le Z_{t+d}$, where $d \ge 1$, we take the difference of the readings in two consecutive days as the water consumption of the second day when the difference of the two consecutive days is one. When the difference of the two consecutive days is more than one, *i.e.*, there exist missing readings in certain days, we impute the missing usage values on those days by using $(Z_t - Z_{t-d})/d$. In summary, the water usage, Y_t on day t is calculated as $Y_t = Z_t - Z_{t-1}$, when Z_{t-1} is available; otherwise $Y_t = \ldots = Y_{t-d+1} = (Z_t - Z_{t-d})/d$.

In addition to the key AMR database, the other data sources are also helpful with understanding the usage patterns, causes of anomalies, and predicting water demand in the future. Customer attribute data include information of customer types, such as residential, commercial, federal and multiple family accounts. Obviously, the demographical information has significant impact on water usage patterns across seasons. Moreover, weather information, such as temperature and rainfall amount, also influences the customers' water consumption behaviors. We collect weather data at the customer service area during the same time period, January 1st, 2008 to June 30th, 2011, from National Climate Data Center [5]. In Section III, we can see that temperature and rainfall have different impacts on water usage patterns for commercial and residential customers.

Additional data sources, such as meter attributes including meter age and materials and customer household size, can be used to explain the malfunctioning meters and provide more accurate water demand forecasting. Depending on the availability of the data sources, our toolkit is flexible to incorporate additional information.

C. Statistical Modeling

The predictive analytical tools include multiple functionalities to perform different analysis. To integrate various functions for smart meter management, we focus on the following three aspects:

a) *Seasonality pattern*. Water consumption of customers present seasonal cycles, such as high consumption in summer and low in winter. There is also a strong 7-day cycles in the data, This 7-day cycle tells us, for example, this Monday's water consumption is more correlated to last Monday's water consumption than other days of the week. For many residential accounts, Saturday's water consumption is usually higher than the other days' consumption amount.

b) *Level shift factor.* In some situations, we observed customers' water consumption level changed at some time point, either due to the change of customer's individual characteristic, such as the change of number of household people or due to the behavior of water theft or bypass. This type of level shift change due to water theft is the water usage anomaly that we are interested in and it can be identified through water usage analysis. The level shift factor is used in the model to capture the sudden change pattern.

c) *Overall trend.* On top of seasonal cycle and level shit, the overall trend over time is estimated. It indicates meter silting problem when the overall trend is significantly negative.

After some careful explorations, for each meter, we decide to fit the following model

$$\begin{aligned} y(t) = {} & \mu + \beta_1 T_1(t) + \beta_2 T_2(t) \\ & + \alpha_1 W_1(t) + \cdots + \alpha_7 W_7(t) \\ & + \gamma_1 M_1(t) + \cdots + \gamma_{12} M_{12}(t) \\ & + \delta_1 \mathrm{Temp}(t) + \delta_2 \mathrm{Temp}(t-1) \\ & + \rho_1 \mathrm{Rain}(t) + \rho_2 \mathrm{Rain}(t-1) \\ & + \mathrm{outliers} + \mathrm{levelshifts} \\ & + N(t), \end{aligned} \tag{1}$$

where t is time index with $t = 1, 2, \ldots$, representing 1st day, 2nd day, etc. The $y(t)$ is the daily water usage for a meter. The $T_1(t)$ & $T_2(t)$ are two trend series, $T_2(t)$ is for the most recent months $t > t^*$, where t^* is the threshold indicator, for example 6 months, which we choose to split the two trends $T_1(t)$ & $T_2(t)$. Then $T_2(t)$ refers to the most recent 6 months, and $T_1(t)$ refers to the months before that. The reason we use two terms, $T_1(t)$ & $T_2(t)$, to model the trends is to differentiate the trends of the most recent 6 months (January 1st, 2011 to June 30th, 2011) and the previous three years (January 1st, 2008 to December 31st, 2010). From the perspectives of business operations and revenue, the water utility companies care more about the most recent performance of the meters. $W_1, \cdots, W_7$ are the weekday indicators from Monday to Sunday, and $M_1, \cdots, M_{12}$ are the month indicators from January to December. They are formulated as

$$T_1(t) = \begin{cases} t & t \le t^* \\ t^* & t > t^* \end{cases}$$

$$T_2(t) = \begin{cases} t^* & t \le t^* \\ t & t > t^* \end{cases}$$

$$W_1(t) = \begin{cases} 1 & t \text{ in Monday} \\ 0 & \text{otherwise} \end{cases}$$

...

$$W_7(t) = \begin{cases} 1 & t \text{ in Sunday} \\ 0 & \text{otherwise} \end{cases}$$

$$M_1(t) = \begin{cases} 1 & t \text{ in Janurary} \\ 0 & \text{otherwise} \end{cases}$$

...

$$M_{12}(t) = \begin{cases} 1 & t \text{ in December} \\ 0 & \text{otherwise} \end{cases}.$$

Please note that there are one redundant week indicator series and one redundant month indicator series, these can be easily taken care by dropping one series, say W_7 and M_{12}. The *Temp(t)* and *Rain(t)* are the daily temperature and rainfall series in the area we study.

$N(t)$ is the residual term after removing the trends, weekly and monthly seasonality, weather influences, etc. $N(t)$ follows the time series AutoRegressive-Moving Average model, ARMA(2,2)(1,1), i.e.,

$$(1-\varphi_1 B-\varphi_2 B^2)(1-\Psi_1 B^7)N(t)=(1-\theta_1 B-\theta_2 B^2)(1-\Theta_1 B^7)a(t)$$

where *a(t)* is a white noise series, φ_1,φ_2 are non-seasonal autoregressive (AR) parameters, θ_1 , θ_2 are non-seasonal moving average (MA) parameters, and Ψ_1 , Θ_1 are seasonal AR and MA parameters, B is back-shift operator such that $B^l X(t)=X(t-l)$. *N(t)* captures the temporal correlation among the daily water usage.

For outliers (isolated outlying points) and level-shifts, there may be multiple of them in one water usage series. Their locations and sizes will be automatically detected.

The model in (1) is a special case of transfer function models [6]. This model can be easily extended to include other external series that may affect water usage. The model was fit using IBM SPSS's time series analysis software [7].

III. Application

We demonstrate our predictive analytics toolkit using the application of the water utility customer. They have about AMR data from about 113,000 accounts. We used the readings from January 2008 to June 2011.

First of all, we visualize the overall pattern and trend of averaged daily water usage of commercial (the top panel) and residential (the bottom panel) customers from January 1st, 2008 to June 30th, 2011 in Figure 3. The unit of daily water usage is cubic feet (ccf). Even though the water utility customer has additional account types, such as Federal and multiple family accounts, commercial and residential customers account for a large proportion of accounts and are typical types for water usage analysis. Thus, we choose these two types for illustration throughout the paper. In Figure 3, an obvious seasonality pattern can be observed from both types of customers, i.e., higher water consumption in summers and lower in winters.

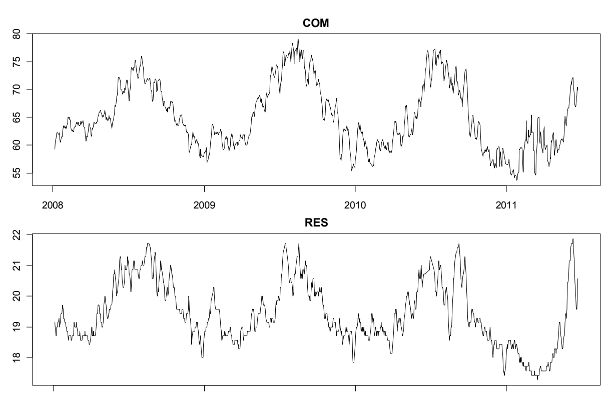

Figure 3: Averaged daily water usage of commercial (top panel) and residential (bottom panel) customers from January 1st, 2008 to June 30th, 2011.

To justify the temporal cycle components in our time series model (1), we show the monthly and weekly seasonality cycles of water usage in Figure 4. The analysis is based on the averaged daily water usage of commercial and residential customers as shown in Figure 3. We aggregate the daily water usage into each month and each day of a week over the three and a half year. The left two panels of Figure 4 show the seasonal patterns of water usages for commercial (top panel) and residential (bottom panel) customers by months. These two types of accounts behave similarly in terms of water consumption cycle over seasons. The right two panels plot the weekly patterns of water usages of commercial (top panel) and residential (bottom panel) customers starting from Sunday labeled as 0 to Saturday labeled as 6. Commercial and residential customers have very different weekly patterns. Residential customers consume more water over the weekend compared to commercial customers who consume more water over weekdays, especially from Wednesday to Friday.

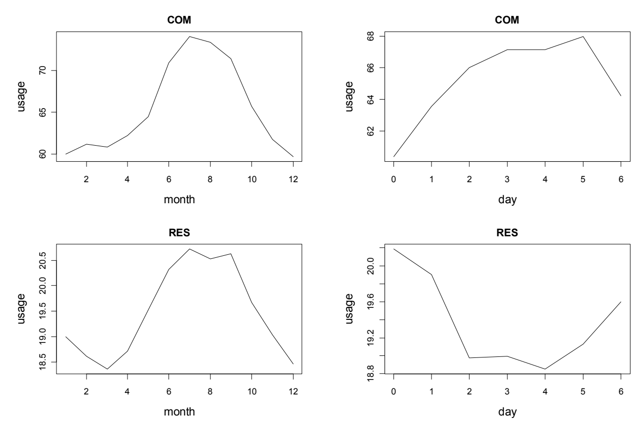

Figure 4: Exploration of monthly (left two panels) and weekly (right two panels) patterns for commercial (top two panel) and residential (bottom two panel) accounts.

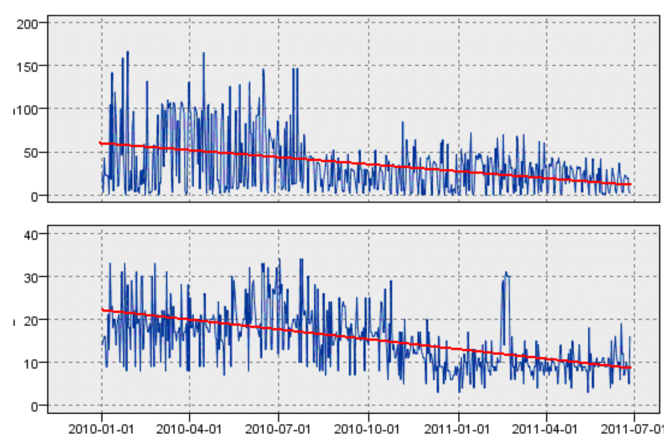

Figure 5: Examples of meter silting.

After exploring the data, we chose the model (1) and perform the analysis which serves the functions of anomaly detection and water demand forecasting. Figure 5 presents two examples of meters with siltation problems. Daily water consumption from January 1st, 2010 to June 30th, 2011 is plotted. A gradually decreasing trend can be observed from both of the trace plots indicated by the red straight lines. Currently, our water utility customer is working on the field pilot to test the siltation detection results for large meters with meter size larger than 2 inches. The accuracy rate is 86%. They are pleased with being able to focus on their crews on investigating meters that have a high probability of having problems to minimize unnecessary truck rolls. Customer Service has back-billed a few customers given the results and intends to do more.

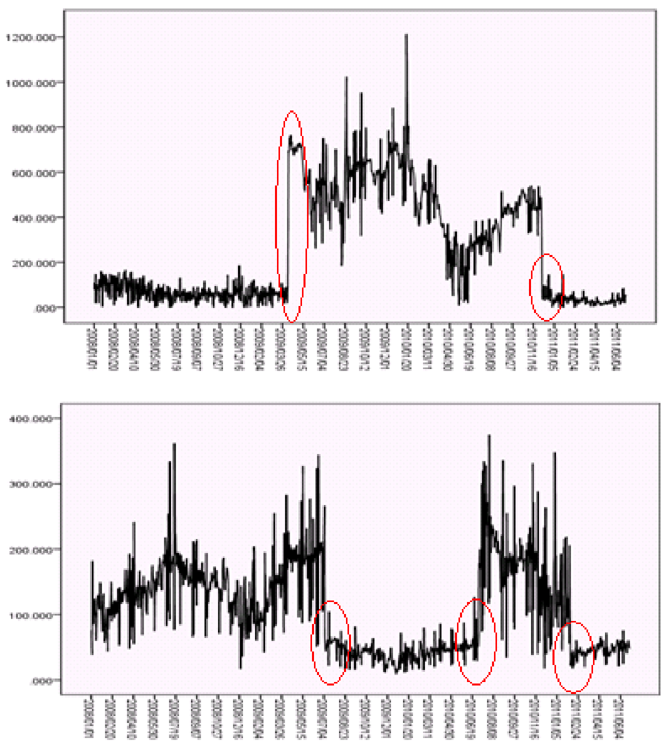

Figure 6: Two examples of level shifts.

Figure 6 presents two examples of level shifts. In water usage applications, the level shifts indicate a sudden change of water consumption patterns which can be caused by water theft or leaking problem. The red circled areas are identified by our model including level shift indicators.

Figure 7 shows two examples of water usage forecasting. The red lines are observed water usage data; the blue lines are forecasted values; the two purples dashed lines are the upper and lower prediction interval respectively. For each account, given its historical water usage data to the current day t, we use model (1) to predict water consumption for the next day, $t+1$. Then we shift the data window one day to include the reading at day $t+1$ in history and predict water demand for the following day, $t+2$. Given this procedure, the water demand on each day is forecasted. The water demand prediction (blue curve) lines up with the actual readings very well. This prediction method can be also used to provide prediction for segmented customers, such as by customer types, commercial and residential, and at different temporal resolution, such as weekly or monthly basis.

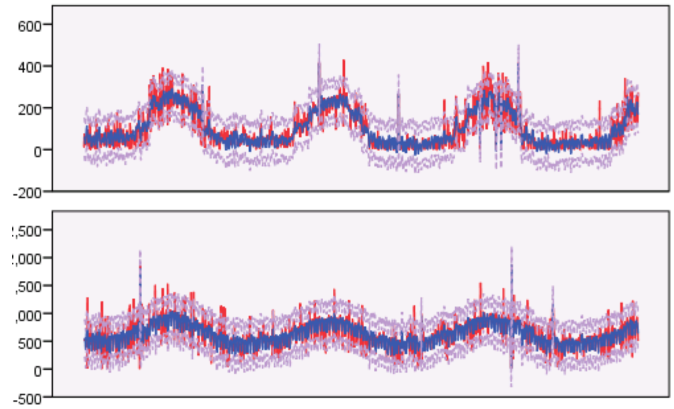

Figure 7: Two examples of water usage forecasting.

In addition to anomaly detection and water demand forecasting, we explore the weather impact on water consumption. The weather factors we consider include daily temperature and rainfall. Since temperature has a seasonal pattern which is consistent with that of the water consumption, we need to remove the confounding concerns. As shown in our model (1), we incorporate monthly indicators which account for the seasonal impacts. Then we can study the "pure" weather impact. The model fitting shows that the temperature on current day has positive impact on the water usage of the same day for both residential and commercial customers. Interestingly, previous two days' rainfall has negative impact on the current day's water consumption for residential customers compared to no impact on commercial customers.

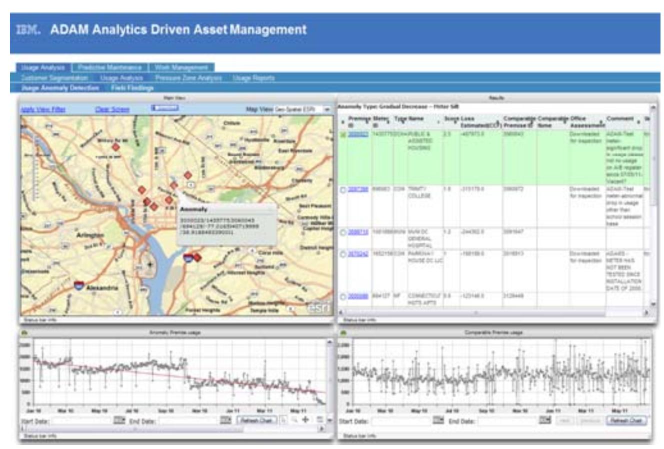

Figure 8: snapshot of ADAM predictive analytics toolkit.

Figure 8 shows a snapshot of the predictive analytics toolkit integrated in the IBM Analytics Driven Asset Management (ADAM) product. The top menu gives the various functionalities of asset management tools. One of the functions under the tab of Usage Analysis is Usage Anomaly Detection. Figure 8 only presents an example of meter silting problem detection. On the top left panel of the screen is the distribution of the certain detected meters. The top right panel shows the list of the detected meters. When one meter is selected, the bottom left panel will show the water usage trace plot over time. Compared to the selected anomaly meter, a normal meter without being identified as anomaly is shown on the bottom right panel.

IV. DISCUSSION

This predictive usage analytics toolkit provides a general tool in the applications of smart water management. Its multiple functionalities serve the various aspects in smart water management including field testing of malfunctioning meters, revenue forecasting and resource planning and scheduling. This analytics tool can be integrated with an optimization tool to further complete the business solution in the areas of planning and scheduling.

ACKNOWLEDGMENT

We acknowledge the support from DC Water and IBM ADAM team.

REFERENCES

[1] Accenture, "Ten Leading Practices for Smart Grid Analytics". [Online] Available: http://www.accenture.com/us-en/Pages/insight-ten-leading-practices-smart-grid-analytics.aspx

[2] http://www-03.ibm.com/press/us/en/pressrelease/28859.wss

[3] http://www.dcwater.com/site_archive/news/press_release419.cfm

[4] A. Hampapur, H. Cao, A. Davenport, W. S. Dong, D. Fenhagen, R. S. Feris, G. Goldszmidt, Z. B. Jiang, J. Kalagnanam, T. Kumar, H. Li, X. Liu, S. Mahatma, S. Pankanti, D. Pelleg, W. Sun, M. Taylor, C. H. Tian, S. Wasserkrug, L. Xie, M. Lodhi, C. Kiely, K. Butturff, L. Desjardins, "Analytics-Driven Asset Management", IBM Journal of Research & Development, vol. 55, no. 1/2, JANUARY/MARCH 2011.

[5] National Climate Data Center [Online]. Available: http://www.ncdc.noaa.gov/oa/ncdc.html

[6] G. Box, G. Jenkins and G. Reinsel, "Time Series Analysis: Forecasting and Control", 3rd ed., Prentice Hall, Englewood Cliffs, NJ. 1994.

[7] IBM SPSS software: http://www-01.ibm.com/software/analytics/spss/