Predictive Source Control: Application of Advanced Numerical Methods and Surrogate Monitoring for Microconstituents in Water Reclamation

Christopher Stacklin, P.E.[1*]

[1] Orange County Sanitation District, 10844 Ellis Avenue, Fountain Valley, CA 92708-7018
*Email: cstacklin@ocsd.com.

ABSTRACT

Having the water reclamation capacity of 70 million gallons per day (MGD) and expandable up to 130 MGD, the Ground Water Replenishment (GWR) System is the largest water purification and reuse project of its kind in the world. With the advent of groundwater replenishment and reuse, a new paradigm of source control has started at the Orange County Sanitation District (OCSD). Traditionally, Source Control was primarily based on enforcement and compliance of the EPA's National Pretreatment and NPDES Programs. At OCSD, Source Control has expanded its role to monitoring and controlling microconstituents from both point and nonpoint sources using two innovative methods: 1) Predictive Modeling and 2) Real-Time Modeling.

As part of its expanded program to assure that the water produced by the GWR System is of the highest quality, Source Control has developed key programs for both predictive and real-time monitoring of microconstituents in source water received by the facilities. Information from predictive and real-time monitoring can be used to alert facility operations of abnormalities such as microconstituent concentration spikes, growth trends, or reductions. The information is also used as triggers for implementation of various point and nonpoint source control measures to reduce or smooth concentration spikes.

Predictive Modeling
Predictive modeling is based on a stochastic, time-series approach using Kalman filters and autoregressive integrated moving average methods such as Box-Jenkins. These models are applied to microconstituent analytical data of 1) the source stream recognizing that there is a high potential of outliers and noise due to matrix interference and signal suppression; and 2) intermediate streams where noise due to matrix interference and signal suppression are diminished but facility dynamics come into play.

Unlike conventional constituents where levels of concentrations are established or exhibit discernable patterns, microconstituents may follow consumer trends that may be influenced by competing new products, product replacement and substitution, health advisories, or consumer news. Historical (past) data can potentially bias the results if it is not discarded after new consumer trends emerge. Therefore, a recursive approach is necessary where the output is used to condition the estimate and past data is discarded.

Results of the stochastic time series model are compared with economic indicators such as pharmaceutical sales estimates or usage statistics, and heuristics based on expert opinion and

product literature. In this manner, the predictive model is tempered with real-life judgment.

Real-Time Modeling
Real-time modeling uses surrogates and analytical equipment to identify the magnitude of concentration and upstream locations of point and nonpoint sources or in-plant. Statistical methods are employed that test for correlation of the surrogate to speciated analytical data. If the surrogate concentration increases, it may indicate an increase in concentration of a family of microconstituents. The use of surrogates, in effect, reduces the time and cost of tracking individual microconstituents. Control and action levels for surrogates are set by a mass balance around the facilities using regulatory limits, standards, or guidelines and removal efficiencies derived analytical values of microconstituents and stream flow rates.

Predictive and real-time models for microconstituents will be developed based on analytical results and flows from the GWR System and Reclamation Plant No. 1. The results will be discussed relative to practical (day-to-day) implementation and use.

KEYWORDS: Microconstituents, Groundwater Replenishment, Surrogates, Predictive Modeling, Real-Time Modeling, Kalman Filer, Robust Regression on Order, Dynamics.

THE CHALLENGE – THE NEED FOR REAL-TIME DETECTION

Microconstituent Monitoring Logistics and Frequency
Microconstituents, also known as emerging pollutants of concern, or micropollutants, etc. occur at relatively low concentrations, tend to be both ubiquitous and persistent in the environment, and have known to unknown toxicity effects.

At water reclamation facilities, there is an emphasis on characterizing microconstituents in the influent streams received by the facilities for the purpose of meeting operational objectives and implementing effective source control of microconstituents being discharged into the wastewater.

Based on today's analytical technology, monitoring frequency of microconstituents is a balance between regulatory requirements, logistics, and economics. Sampling for microconstituents is a batchwise versus continuous process and consists of collecting a representative sample, then analyzing it, and reporting results.

Due to practicality and economics, microconstituent sampling is typically taken infrequently whether it is on a monthly, quarterly, or annual basis. This batchwise mode leaves data gaps in the time in-between sampling when there are no samples taken and therefore no analytical results reflecting microconstituent concentrations entering the treatment facilities.

The time from sample collection to reporting results can also be significant. At OCSD, the average turnaround time for analytical data is 36 days from the time that a sample is collected to the time that analytical data is disseminated, although there are ways to expedite turnaround

times.

Implications of System Dynamics

In consideration of the sampling frequency and turnaround time, if a significantly high microconstituent concentration is detected in a sample, what is the available response time to adjust operations or implement source control measures?

The response time can be defined as the time that a microconstituent would take from its point of discharge, through the collection system, then to the start of treatment. As an example, a microconstituent would flow through the: 1) regional collection system, through 2) Reclamation Plant No. 1 treatment processes, then through 3) microfiltration before it reaches reverse osmosis and advanced oxidation where it would be removed or destroyed.

Hydraulic detention time in the regional collection system within Orange County can range from instantaneous to about 8 hours, averaging 4 hours. Next, wastewater from the collection system is received by Reclamation Plant No. 1. At Reclamation Plant No. 1, the hydraulic detention time ranges from 4 hours to 8 hours, depending on the diurnal flow received by the plant. Diurnal flow is the daily variation in flow received by the treatment facilities. The change in flow rate received by the treatment facilities is influenced by the water use habits of 2.5 million residents in Orange County. During the day, people work and tend to use more water than during the night when they are sleeping. Also, heavy industrial water users operate in the day. The flow variation at OCSD typically ranges from a 116 MGD peak at about 8:00 am to a 40 MGD trough at about 4:00am. Finally, hydraulic detention time across the GWR System upstream of reverse osmosis is 40 minutes based on an estimated 2 million gallon detention capacity at the current rated capacity of 70 MGD.

Thus, the average time it takes for a chemical constituent to reach the treatment facilities is 8½ hours versus the analytical processing time of 36 days. If sampling is done on a daily or weekly basis, then the time lapse that a potential problem can be detected ranges from 15½ hours to up to 7 days.

A Logistical Conundrum

Say that 1,4-dioxine is accidentally being discharged into the collection system due from a small leak which is undetected at the source. If 15.5 hours up to seven days elapse before the concentration spike is detected, and assuming that it is at the action level of 3 parts per billion, then the potential additional mass reaching treatment is 1 pound to 7.3 pounds, respectively.

If the treatment system is operating at its design capacity, then there is risk of breakthrough. Although 7.3 pounds is very small compared with a reservoir which may contain 200,000 acre feet of water, drinking water quality regulations typically are concentration -based rather than mass -based.

SHIFTING FROM REACTIVE TO REAL-TIME, PREDICTIVE SOURCE CONTROL

Traditional sampling is batchwise and requires turnaround time from sample collection to analysis that typically exceeds the hydraulic detention time of the water treatment technology and collection system. Recent technology reduces turnaround time from days and weeks to almost instantaneous and continuous measurement.

Analysis in Real-Time

New analytical equipment is available which allows on-line, in stream and continuous sampling for parameters including total organic carbon (TOC), conductivity, pH and temperature. TOC is an indicator of the presence of organics, although it lacks speciated data to define exactly what constituent was detected which would narrow down the sources of discharge. Conductivity is an indicator of the presence of salts such as metals. pH indicates presence of acids or bases. Temperature measurement is required for pH correction. These real-time and continuous parameters help to quickly identify new trends such as an increase in concentration or narrow down the source of the discharge.

Conceptually, these parameters which detect in real-time function as surrogate indicators for classes of chemical compounds. A surrogate is something to take the place for another thing, in this case, TOC will take the place of speciated analytical data for organic compounds. An indicator is a measure of difference between a reference or standard and the observation, such as an organic chemical grouping versus TOC which serves as a standard.

A challenge to this technology is that it is less effective in a wastewater matrix. A wastewater matrix is both corrosive and fouling, requiring constant maintenance. Also, the wastewater matrix tends to cause analytical interference and signal suppression.

Analytical equipment can be deployed in strategies areas of the treatment technology unit processes and collection system to reduce the time for detection. Further, analyzing key parameters which may serve as surrogates will also reduce the resources and cost required to maintain such programs. For example, TOC analysis can be routinely used to detect significant changes in organics concentration instead of running a detailed characterization for volatiles and semi volatiles, and base/neutral/acid organic compounds. If an excursion is observed in the TOC analysis, then a detailed characterization can be performed to identify significant compounds.

Characterization is paramount to determining the root cause of the excursion. When analytical concentration is combined with flow data, the mass flux of a constituent entering the system can be assessed. Also, by deploying the system in strategic areas and utilizing a geographic information system (GIS), the root cause may be quickly isolated to a local area.

Forecasting Micropollutant Concentrations

Since real-time monitoring can detect general changes in parameters instantaneously, recalling that the dynamics of the regional collection system ranges from instantaneous near the treatment

facilities to an average of 4 hours, real-time modeling is still more of a reactive element versus pro active. Therefore, to have a pro-active program, making forecasts from the analytical concentration data is necessary.

Advanced numerical methods have been used as a basis for forecasting in the past. Challenges with respect to predicting concentrations of microconstituents are: 1) the presence of interference and noise in the data; 2) the impact of left-censored data on the prediction; 3) outliers and 4) the impact of past data that is no longer relevant. Following is the method development for the predictive source control model deployed at OCSD.

METHODOLOGY AND DESIGN BASIS DEVELOPMENT

Numerical methods used to model microconstituent concentrations were carefully chosen based on defining a detailed design basis.

Describing the Water Reclamation Domain Space

Why model the influent concentration to a wastewater treatment plant (WWTP) as a stochastic time series process?

Univariate Time Series

By definition, a time series is an ordered sequence of observations (analytical measurements) made over a continuing length of time. In the case of WWTP's, influent concentrations of microconstituents such as 1,4-dioxane are periodically and continuously measured. If the sequence of observations is comprised of a single set of numbers, then the time series is univariate. If the sequence of observations is n-dimensional vectors, then it is multivariate, and *n* is the dimensionality of the time series. For example, consider the time series dataset of a single constituent, ammonia in Table 1. The data comprises a univariate time series where $\{z_{k-6}, z_{k-5},..., z_k\}$ = {30.1, 26.4, ... , 30.6}.

Table 1. Ammonia average daily concentration, mg/L.[a]

2/18/2009	2/19/2009	2/20/2009	2/21/2009	2/22/2009	2/23/2009	2/24/2009
30.1	26.4	30.8	30.8	30.5	28.6	30.6

[a] Data is from Phase IV sampling data.

An example of a multivariate time series would be one comprising not just ammonia concentrations, but nitrates, and nitrites concentrations as well. This would be a 3-dimensional time series where for *i* number of microconstituents, a datapoint observation would be represented as, $z_{i,k}$. This paper will develop a univariate time series model, for simplification.

Discrete Stochastic Process

Observations that generated a time series, such as microconstituent concentrations will continue into the future. Future microconstituent concentration values are of interest, and are treated as random. Therefore, to model these values, a model called a stochastic process based upon the time series is used. The word stochastic means random.

In probability theory, a stochastic process is the counterpart to a deterministic process. A deterministic process deals with only one possible outcome at a future time based on an initial condition. In a stochastic process, an outcome may be one of many possible outcomes which can be described by a probability distribution. This means that even if the initial condition such as a past observation of microconstituent concentration is known, there are many possibilities that the future concentration may be at, but some paths are more probable and others less.

A stochastic process is a set of random variables or random vectors ordered with respect to time *t*. If *t* takes on integer values, the process is a discrete-time or discrete process. If it takes on real values, it is a continuous-time or continuous process. Since microconstituent sampling is a batchwise versus continuous process, the domain is therefore a discrete time-series stochastic process.

Time series analysis is the fitting of stochastic processes to the time series. This typically involves statistical analyses, but it is not a straightforward application of statistics. Examples of processes modeled as stochastic time series include stock market and exchange rates, predicting weather, GPS navigation and autopilot, robotic vision, audio and video signals, etc.

Stationarity

Another criterion is if microconstituent concentration values represent a stationary process or non stationary process. A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time or space. As a result, parameters such as the mean and variance, if they exist, also do not change over time or position. For example, analytical matrix interference is stationary, but the sound of an echo is not because it diminished over a period of time.

Examples of discrete-time stationary processes with continuous sample space include autoregressive and moving average processes which are both subsets of the autoregressive moving average model, Markov chains, and the Kalman filter.

To sum up the water reclamation domain space, water quality or more specifically microconstituent concentration can be described as a univariate time series, discrete stochastic process with stationarity. Therefore, a Kalman filter is proposed as a predictive model.

Kalman Filter

What is a Kalman filter? A Kalman filter is a recursive filter that estimates the state of a linear dynamic system from a series of noisy measurements (Welch Greg, et. Al., 2006). Signal noise, for example, is fluctuation and external factors added to a datastream signal received by a detector, such as an analyzer. The receiving device can also be source of more signal noise. Data is corrupted with this noise. Kalman filters are a means to filter noise from data. A good filtering algorithm can filter the noise from a datastream while still retaining useful information. Rudolph E. Kalman is credited in publishing the algorithm and approach in 1960 in his landmark paper, “A New Approach to Linear Filtering and Prediction Problems” (Kalman, R.E. 1960).

Kalman filters were initially used largely for data compression algorithms in weather satellite communications in the early 1960's. Since then, Kalman filters are commonly used as predictors for the stock market and econometrics, robotic navigation and visualization and computer animation, to name a few.

The Kalman filter is ideally suited for stochastic, time series systems or domains, which is why it can be applied to filtering the noise in analytical data of microconstituents (Stacklin, Christopher, 2008)(Stacklin, Christopher and Evangelista, Jerry, 2008). Contributors to noise in the analytical data are variations in analytical methodology, matrix interference and signal suppression. External factors are changes in influent flow rate, accidental or illegal discharges of chemicals, and behavioral factors of dischargers. A potential stumbling block to the application of Kalman filter or any other time series analysis is that the analytical data of microconstituents is left-censored.

Non Detect and Data Analysis

Analytical data frequently are left-censored due to detection limits of laboratory methods. Left-censored means that some of the observations are known only to fall below a censoring point commonly known as a method detection limit or reporting limit, where the concentration is reported as a non detect. A non detect value simply means that the concentration of a microconstituent may be at or below the method detection or reporting limit. It can even be zero. As an example, a non detect value of <2.0 mg/L means that the concentration is at or between zero and 2.0 mg/L. This presents difficulties in statistical analysis of the data.

In the past, people have been dealing with non detect values by simply substituting the non detect with one half or the full value of the method detection or reporting limit (Helsel, Dennis R., 2006b). Since microconstituents by definition are detected at relatively low concentrations, substitution can bias the results significantly (Helsel, Dennis R., 2005c). Fortunately, there are several methods available for dealing with non detect values that will minimize biasing the data.

What are among the best methods for handling non detect data for microconstituent datasets? Estimating descriptive statistical methods are Kaplan-Meier (KM), Robust Regression on Order Statistics (ROS), and Maximum Likelihood Estimate (MLE) (Helsel, Dennis R., 2005a)(Helsel, Dennis R., 2005d) (Singh, Anita and Nocerino, John, 2006). Note that descriptive statistics are distinguished from inductive statistics in that they aim to quantitatively summarize a data set, rather than being used to support statements about the population that the data are thought to represent. Key attributes of these methods are: 1) the KM is a non parametric method and does not yield values for non detects; 2) Robust ROS and MLE are parametric methods; 3) MLE requires larger datasets then KM and Robust ROS for accuracy.

For time series algorithms such as Kalman filters to work properly, non detects must be converted to values, therefore parametric methods are required. Therefore, KM cannot be used, leaving Robust ROS and MLE. Since Kalman filters use small datasets to reinitialize or restart

due to trend changes and microconstituent datasets by nature will tend to be smaller, the method selected for data preprocessing for the Kalman filter is Robust ROS.

Robust Regression on Order Statistics

An ordinary least squares (OLS) regression estimate assumes that the error term has a constant variance or that observations are drawn from identical distributions. If the error term varies with each observation as is typically the case with time-series measurements, then this reflects that the random variables have different variances. Outliers can have the same influence. Therefore the domain is heteroskedastic.

Heteroskedasticity does not cause OLS coefficient estimates to be biased nor inconsistent. However, the variance (and, thus, standard errors) of the coefficients tends to be underestimated, inflating t-scores and sometimes making insignificant variables appear to be statistically significant. To circumvent underestimation, robust regression is required (Helsel, Dennis R., 2005a).

Implementation of Robust ROS

Details of Robust ROS calculations are described in, *Nondetects and Data Analysis: Statistics for Censored Environmental Data* (Helsel, Dennis R., 2005a). The basic steps to implement Robust ROS are to rank, test, and convert the data for lognormal, normal, or root mean square distribution (Shumway, et. Al. 2002); determine the probability of non detects and regress their values from the distribution (Helsel, Dennis R., 2005a). When the values of the nondetects are defined, the dataset can be then processed by the Kalman filter.

Implementation of the Kalman Filter Model

Note that the Kalman filter algorithm herein includes zeroth, first, and second order filters. A zeroth order filter is good for smoothing the data, e.g., removing the noise to discern general trends. The second order filter is better a predicting values. A second order Kalman filter with restart will be even better at predicting values, but will be described in another paper.

The measurement of the variable, z_k represents a measured concentration value of a microconstituent. The Kalman filter for sequential least squares estimating (Sorenson, H.W., 1970) starts with the definition of the minimum mean square estimator,

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H_k\hat{x}_{k|k-1}) \quad (1)$$

Where

$\hat{x}_{k|k}$ = Minimum mean square estimate
$\hat{x}_{k|k-1}$ = Predicted estimate
K_k = Gain matrix
z_k = Measurement of the variable
H_k = Observation matrix
k = Measurement number

The gain matrix is defined by,

$$K_{k|k} = P_{k|k-1}H_k^T\left[H_k P_{k|k-1}H_k^T + R_{k|k}\right]^{-1} \quad (2)$$

Where
$P_{k|k-1}$ = Covariance of error in the predicted estimate
H_k^T = Transpose of the observation matrix
$R_{k|k}$ = Variance of values from the estimate

Covariance of the estimator error or matrix inversion lemma,

$$P_{k|k} = P_{k|k-1} - P_{k|k-1}H_k^T\left[H_k P_{k|k-1}H_k^T + R_{k|k}\right]^{-1}H_k P_{k|k-1} + Q \quad (3)$$

Where
$P_{k|k}$ = Covariance of error in the estimate
Q = Model variance error

The predicted estimate can be found using the transition matrix such that,

$$\hat{x}_{k|k-1} = \emptyset_{k|k-1}\hat{x}_{k-1|k-1} \quad (4)$$

Where
$\emptyset_{k|k-1}$ = Transition matrix
$\hat{x}_{k-1|k-1}$ = Previous best estimate

The transition matrices, $\emptyset_{k|k-1}$ for the zeroth, first and second order filters are,

$$[1] \quad (5)$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

Corresponding observation matrices, H_k for the zeroth, first and second order filters are,

$$[1] \quad (8)$$

$$[1 \quad 0] \quad (9)$$

$$[1 \quad 0 \quad 0] \quad (10)$$

To initialize the estimator, R_k is the median of the variances (bias) of n joint observations. R_k will remain constant for any *k*. The sample variance (bias) of n measurements is,

$$R_k = \sigma^2 = \frac{\sum_{k=1}^{n}(z_k - \bar{x})^2}{n} \quad (11)$$

Where

$$\bar{x} = \frac{\sum_{k=1}^{n} z_k}{n} \quad (12)$$

The previous covariance of error in the estimate

$$P_{k-1|k-1} = R_k X' X \quad (13)$$

Covariance of error of the predicted estimate is,

$$P_{k|k-1} = \emptyset P_{k-1|k-1} \emptyset^T \quad (14)$$

The predicted estimate is,

$$\hat{x}_{k|k-1} = \emptyset_{k|k-1} \hat{x}_{k-1|k-1} \quad (15)$$

After initialization, the full equations may be used.

RESULTS

Robust Regression on Order

Figure 1 shows the results of the application of Robust ROS to the Estriol dataset. Note that Estriol is used for estrogenic hormone therapy. Out of the 35 datapoints, 13 are non detects which represents 13% of the data. The red line shows the regressed non detect values.

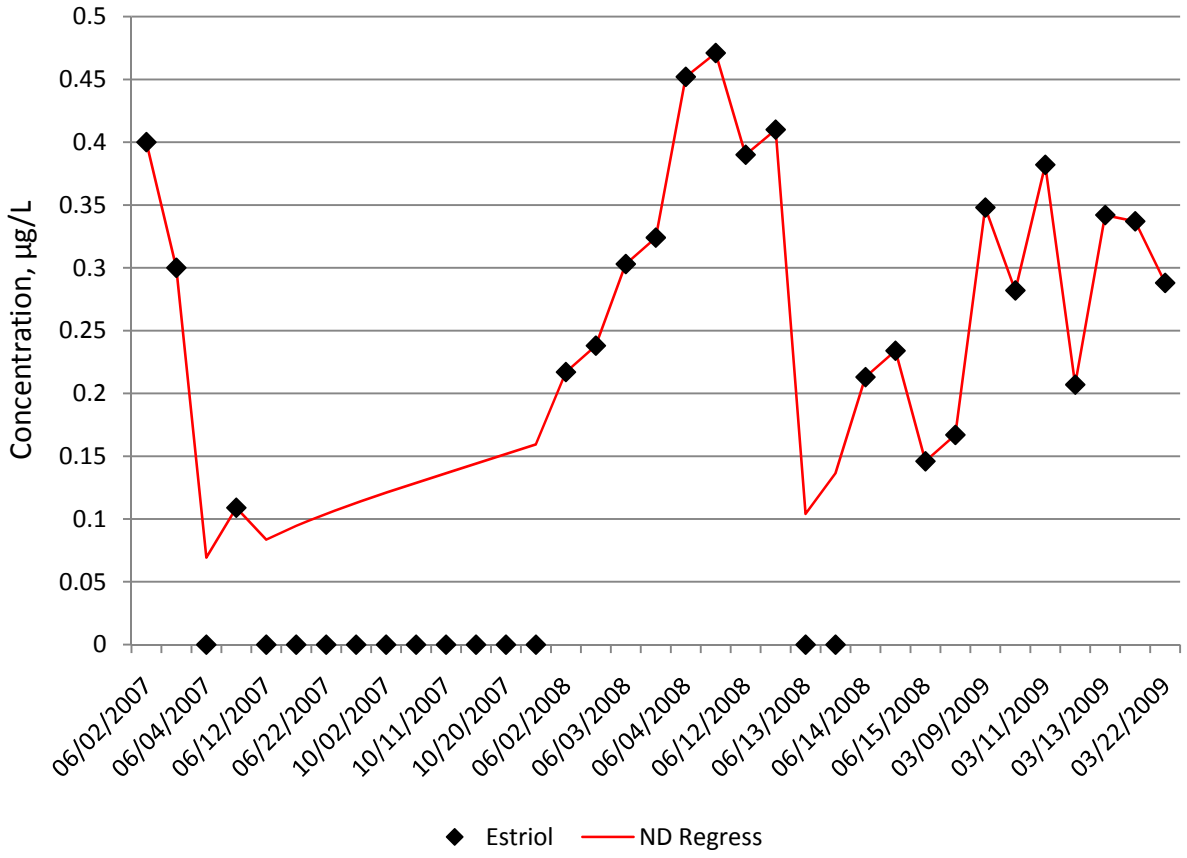

Figure 1. Robust ROS results for Estriol.

Summary statistics are presented in Table 2. Note that substitution of ½ of the method detection limit and maximum likelihood estimate indicate poor results. In this case, the maximum likelihood estimate is off due to the small number of observations. The Kaplan-Meier statistics look good and is useful for small datasets. Unfortunately, Kaplan-Meier is non parametric, e.g., non detect values cannot be derived from the method.

Table 2. Summary statistics for Estriol.

Method	Mean	Std. Deviation	UCL 95%
Detected Values	0.298	0.099	- - -
Substitution ½ Detection Limit	0.191	0.162	0.237
Kaplan-Meier	0.228	0.119	0.263
Robust ROS	0.232	0.118	0.264
Maximum Likelihood Estimate	0.140	0.231	0.206

Figure 2 shows the results of the application of Robust ROS to the Bisphenol A dataset. Note that Bisphenol A is a chemical building block that is used primarily to make polycarbonate plastic and epoxy resins. Out of the 35 datapoints, 6 are non detects which represents 17% of the data. The dotted red line shows the regressed non detect values.

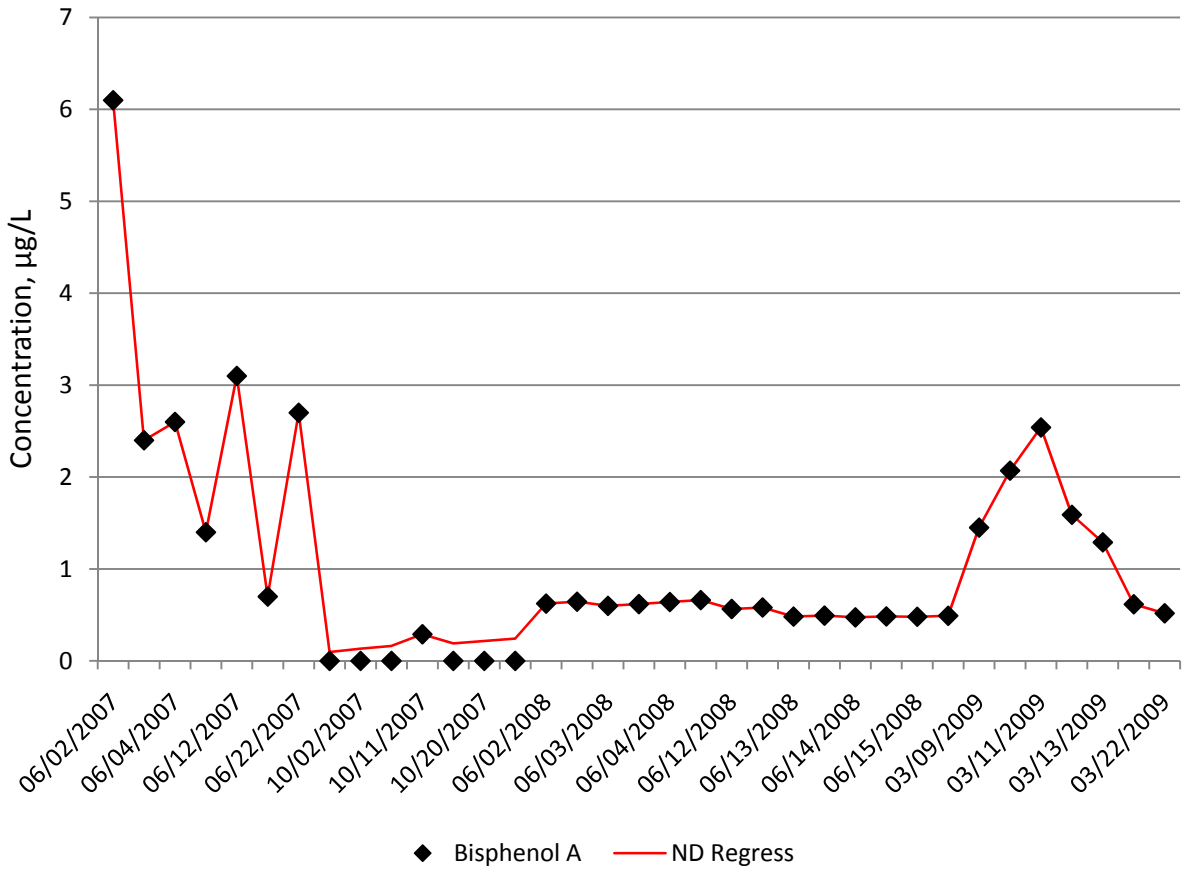

Figure 2. Robust ROS results for Bisphenol A.

Summary statistics are presented in Table 3. Again, substitution of ½ of the method detection limit and maximum likelihood estimate indicate poor results. The Kaplan-Meier is better, followed by Robust ROS statistics.

Table 3. Summary statistics for Bisphenol A.

Method	Mean	Std. Deviation	UCL 95%
Detected Values	1.283	1.248	- - -
Substitution ½ Detection Limit	1.065	1.232	1.417
Kaplan-Meier	1.113	1.177	1.455
Robust ROS	1.093	1.209	1.437
Maximum Likelihood Estimate	0.921	1.400	1.321

Kalman Filter

Kalman filter results are presented below. The Kalman filter results presented are based on a second order filter with no reset. Microconstituent concentrations (observed values) are indicated by a diamond, ◆. Predicted values are indicated by a red line, —.

Estriol is shown in Figure 3 with the residual sum of squares (RSS) value at 0.06. The RSS is a measure of the discrepancy between the data and an estimation model. A small RSS indicates a tight fit of the model to the data. The Kalman filter results are also shows for Bisphenol A with the RSS of 4.75.

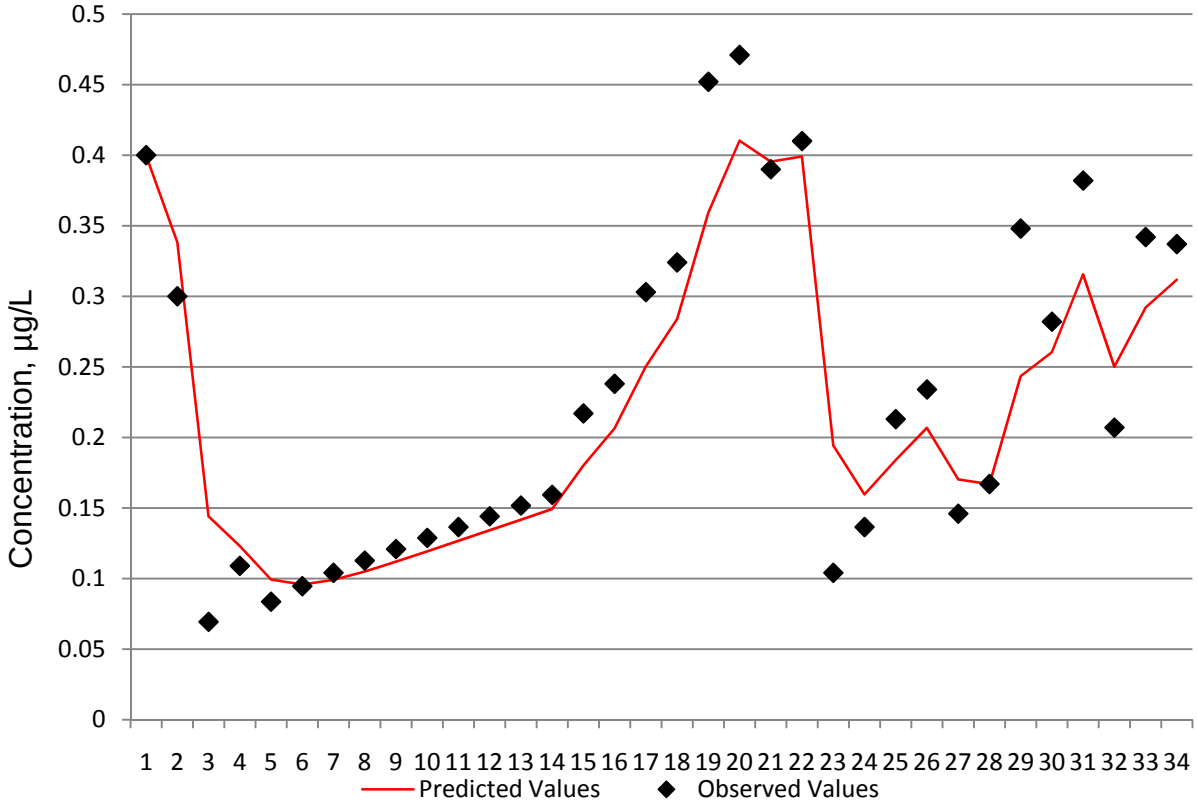

Figure 3. Kalman filter results for Estriol.

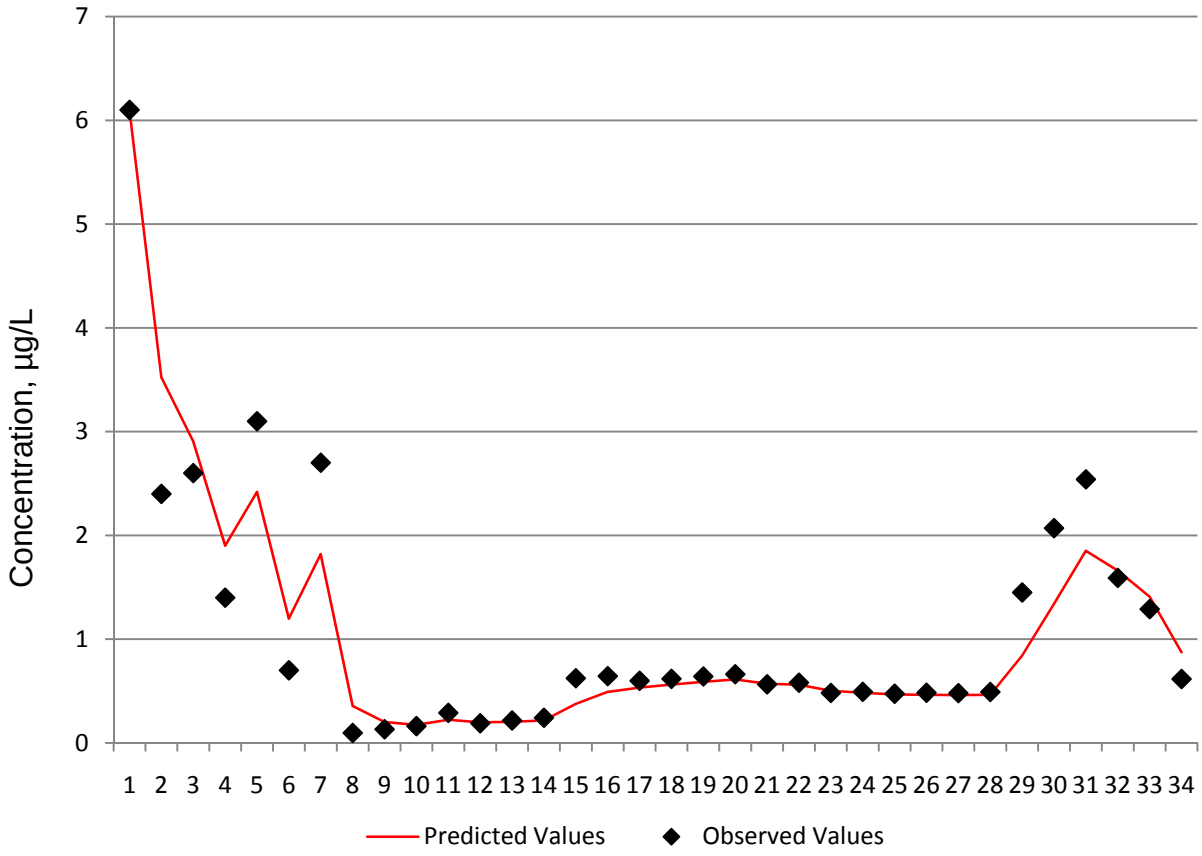

Figure 4. Kalman filter results for Bisphenol A.

Kalman filter versus Moving Average

Figures 5 and 6 compare the Kalman filter results versus two other discrete time series methods, simple moving average (simple MA) and exponential moving average (exponential MA), cousin methods to ARIMA Box-Jenkins. Both methods are very fast and easy to implement in contrast to the Kalman filter. However, the results in Table 4 show that the Kalman filter is a much better fit in both cases.

Table 4. Comparison of RSS values of time series methods.

Microconstituent	Simple MA	Exponential MA	Kalman Filter
Estriol	0.33	0.20	0.05
Bisphenol A	12.17	12.37	1.58

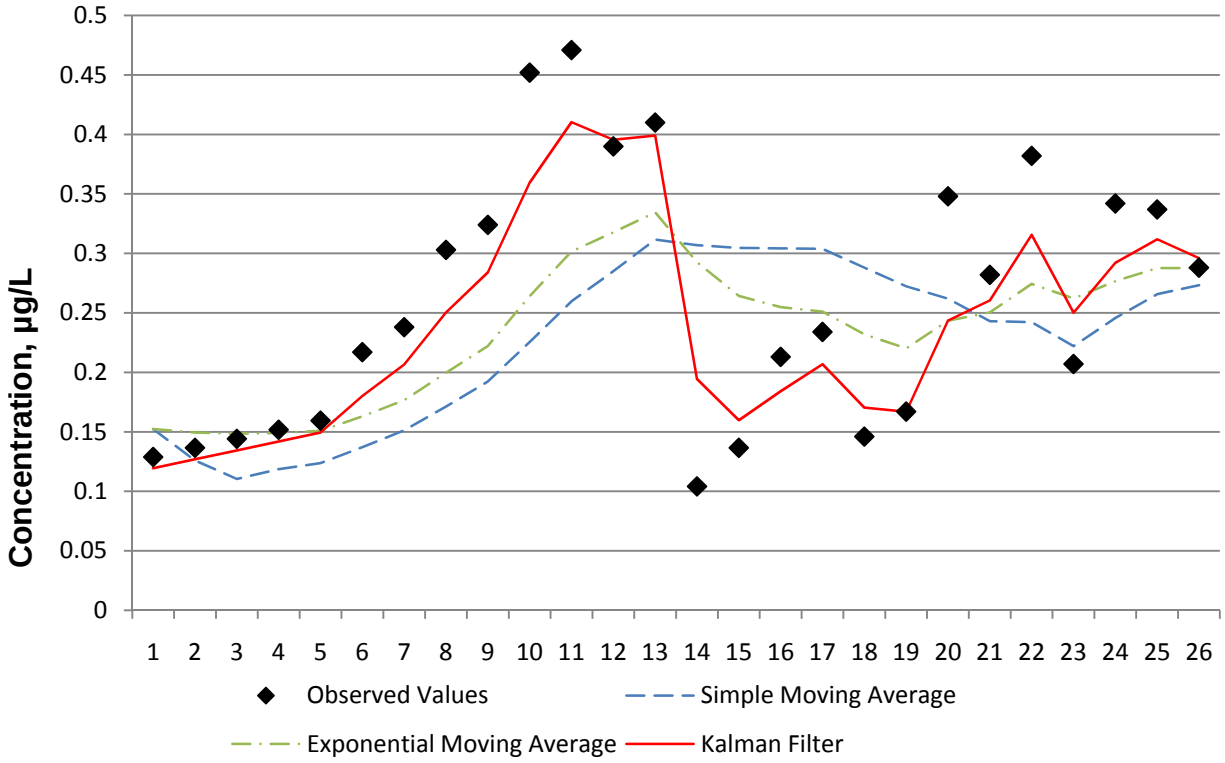

Figure 5. Kalman filter results for Estriol.

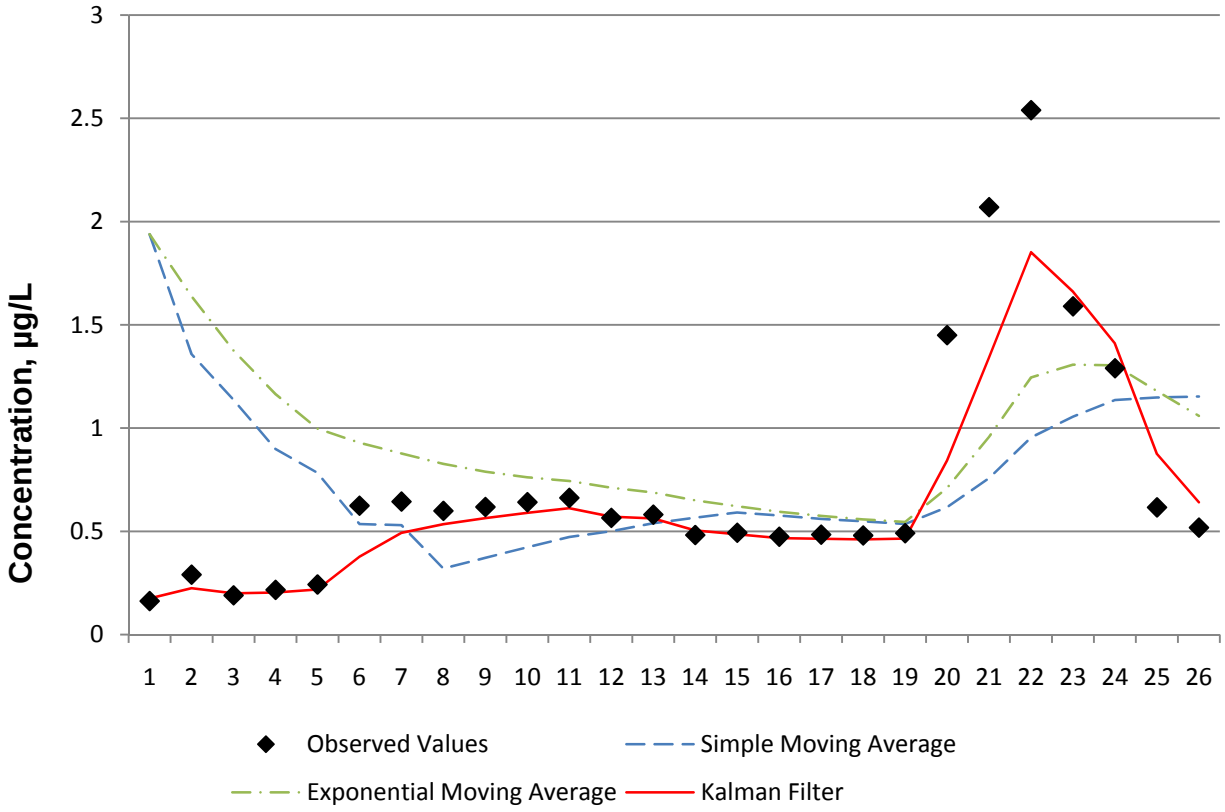

Figure 6. Kalman filter results for Bisphenol A.

TOC as a Surrogate Indicator

It was originally proposed that TOC be used as a surrogate indicator for microconstituents. However, the concentration of TOC at the influent to Reclamation Plant No. 1 averages 80 mg/L. The high TOC concentration relative to the concentration of a microconstituent which may be several orders of magnitude lower than TOC (in the µg/L to ng/L range) would make discerning any relevance very difficult. For example, if the average microconstituent concentration was 50 µg/L, then a microconstituent would constitute about six onehundreths of a percent of the TOC.

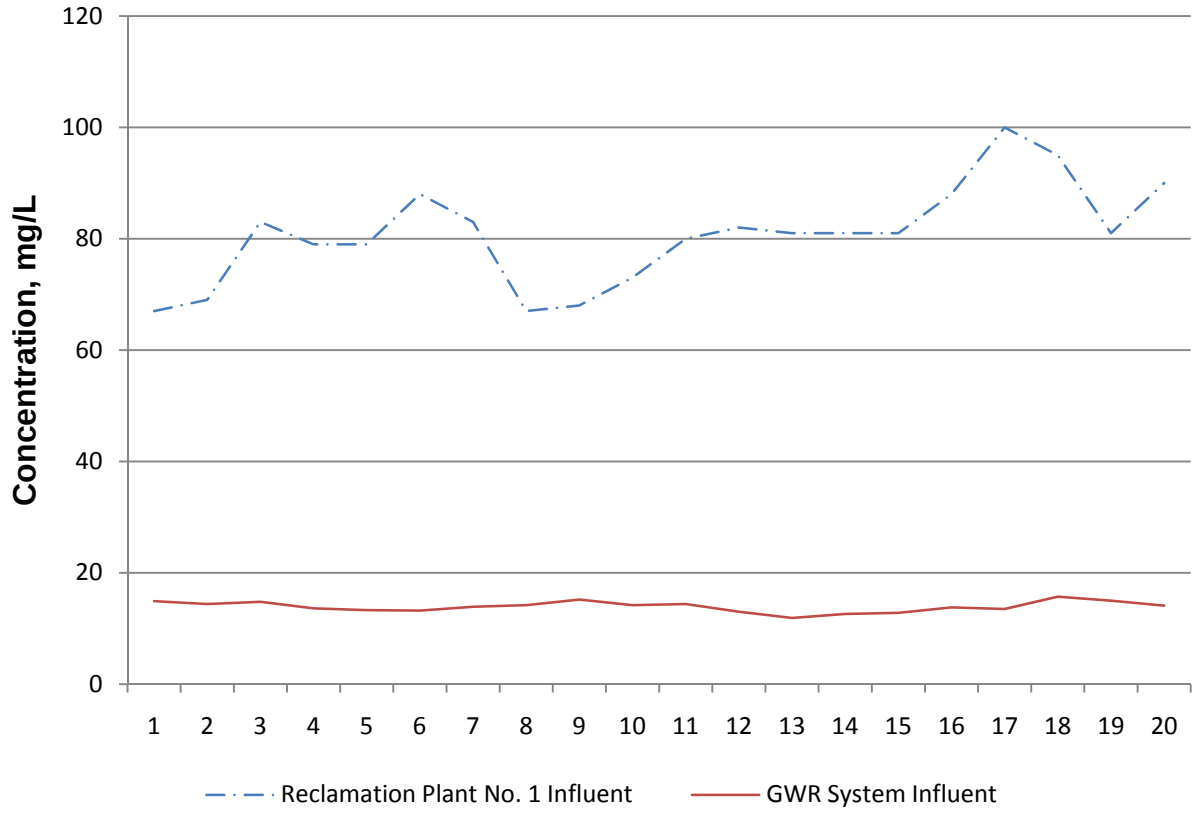

Figure 7. TOC at Reclamation Plant No. 1 versus GWR System influent.

In contrast, the TOC level of the secondary effluent to the GWR System is much lower and averages below 20 mg/L. In this case, an average microconstituent would constitute about one quarter of a percent of the TOC. The wastewater matrix is also much cleaner. Therefore, it may be possible to correlate microconstituent values with TOC based on downstream locations such as immediately upstream of reverse osmosis.

CONCLUSIONS AND DISCUSSIONS

Real-Time Monitoring

While real-time monitoring has a distinct advantage over batchwise sampling, the advantage

gained is limited to the dynamics of the collection system and the treatment system which may average only 8 ½ hours. When real-time modeling is coupled with Kalman filters, the predictive model can extend upward to a week or more, depending on discharge habits.

It may be possible to use TOC as a surrogate indicator, e.g., correlate TOC to microconstituent values, but immediately upstream of reverse osmosis and advanced oxidation processes and downstream of primary and secondary treatment. Placing in-stream TOC monitoring at this location allows for a cleaner matrix to analyze and the TOC value is proportionally closer to the sum of the microconstituent values.

TOC values can be collected in real-time. By converting the TOC continuous telemetry data to discrete data, a Kalman filter can be used to predict future concentrations. If the TOC concentration is predicted to rise significantly, then the TOC is required at the influent to the plant to determine a mass-based removal efficiency. If the removal efficiency is negative, then the TOC is being added or created inside of the treatment processes. If the removal efficiency is positive, then the TOC is being added from the collection system.

TOC analyzers coupled with Kalman filters can be located on the regional trunk lines and pump stations in strategic locations allowing geographic localization of the potential cause of the fluctuation. If the TOC concentration is uniformly increased, then the source is ubiquitous. Therefore a general outreach program, or certification program, or product ban may be considered to achieve a reduction. If the source is localized, then a targeted outreach program, certification, or permitting of a point source or commercial market sector may be considered. The TOC results can be integrated with a Geographic Information System and Chemical Inventory Program to localize the source quickly.

Predictive Kalman Filter

Based on the domain space definition, a Kalman filter was proposed and demonstrated as a model to predict future microconstituent concentrations. It is important to note that the predicted concentration must be converted to a mass value using flow rate in order to have real relevance. A mass value establishes the magnitude of the condition. For example, a concentration of a microconstituent which is approaching an action level, say 20 ng/L equates to 12 pounds per day in a 70 MGD treatment facility. Finding a 12 pound per day source within the treatment facility or being discharged into the collection system can be very challenging. Hence the mass rate puts things into perspective.

The Kalman filter predictor should only be used for near-term prediction initially until a thorough understanding of the behavior of the dynamics of the system is gained.

Tempering Prediction Using Heuristics

It is very important to temper any mathematic forecast by rationalization. This can be done by considering economic trends, marketing growth cycle, publicity and most of all, common sense. These are experience-based techniques that help in problem solving and are known as heuristics.

Complex rules can be chained together using a Bayesian decision model to form an assessment or decision (Winkler, Peter, 2000).

Figure 8 is based on a simple, quick and dirty econometric forecast for the discharge of Estriol in Orange County. The econometric model considers economic criteria to determine Estriol demand. The waste mass load of Estriol received at OCSD was allocated on a per capita basis based on the female population in Orange County at the time of sampling. The female population growth forecast in Orange County is factored in as well as the gross domestic product growth (or decrease) and an assessment of the impact of the FDA's position on the benefits of hormone treatment theory drugs.

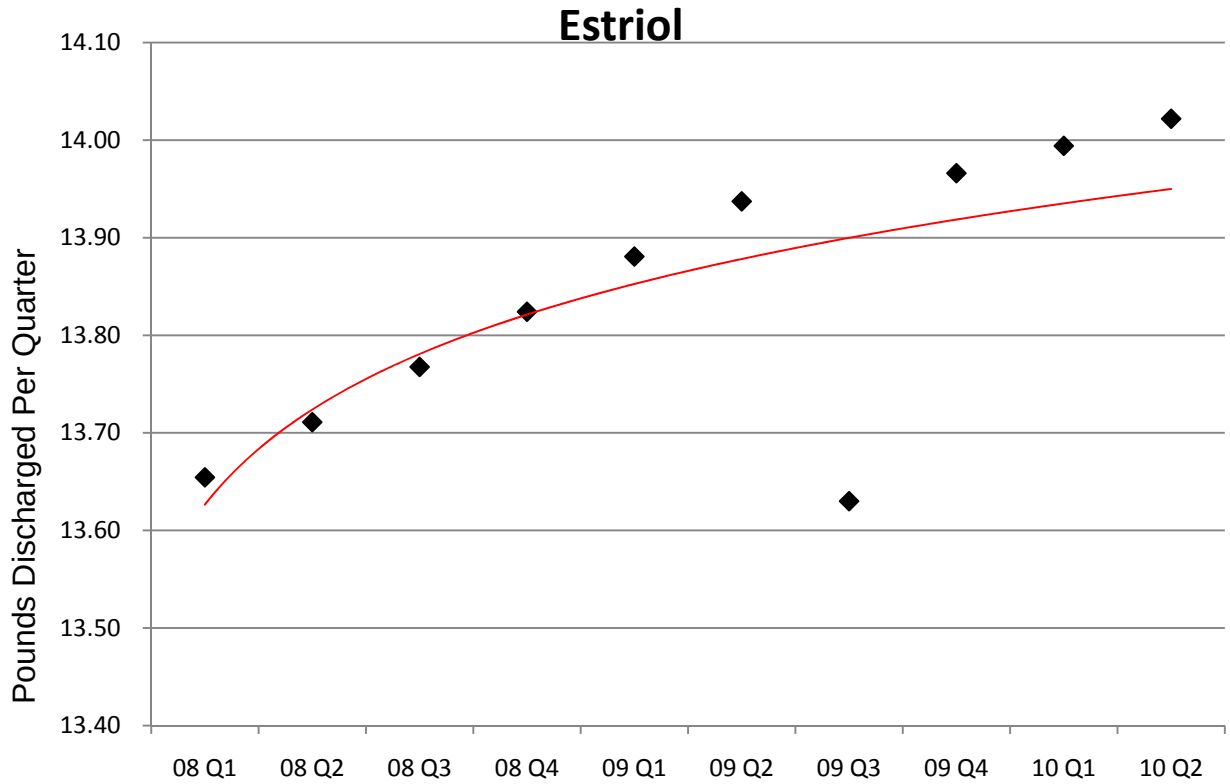

Figure 8. Econometric forecast of Estriol in wastewater discharged in Orange County.

CONCLUSIONS

Water Reclamation Domain Space

Water quality or more specifically microconstituent concentration can be described as a univariate time series, discrete stochastic process with stationarity.

Real-time Modeling

Real-time modeling of water reclamation treatment processes and is possible with recent technology innovations, but the benefits do not allow enough lead time for pro-active operational and source control programs. Real-time modeling can provide a finite amount of lead time

which could range from hours to as much as a day depending on system dynamics.

Predictive Modeling

It was demonstrated that a Kalman filter can be used to provide near-term prediction of microconstituent concentrations ranging from days to weeks depending on the behaviors of dischargers.

Robust ROS, an estimating descriptive statistical method, was used to regress values for non detects in the analytical datasets while minimizing any biasing of the descriptive statistics.

Results from the Kalman filter predictive model can be tempered using an econometric, heuristic model that describes local demographic trends and subjective impacts, e.g., consumer news, regulations, etc. Figure 9 shows the predictive model.

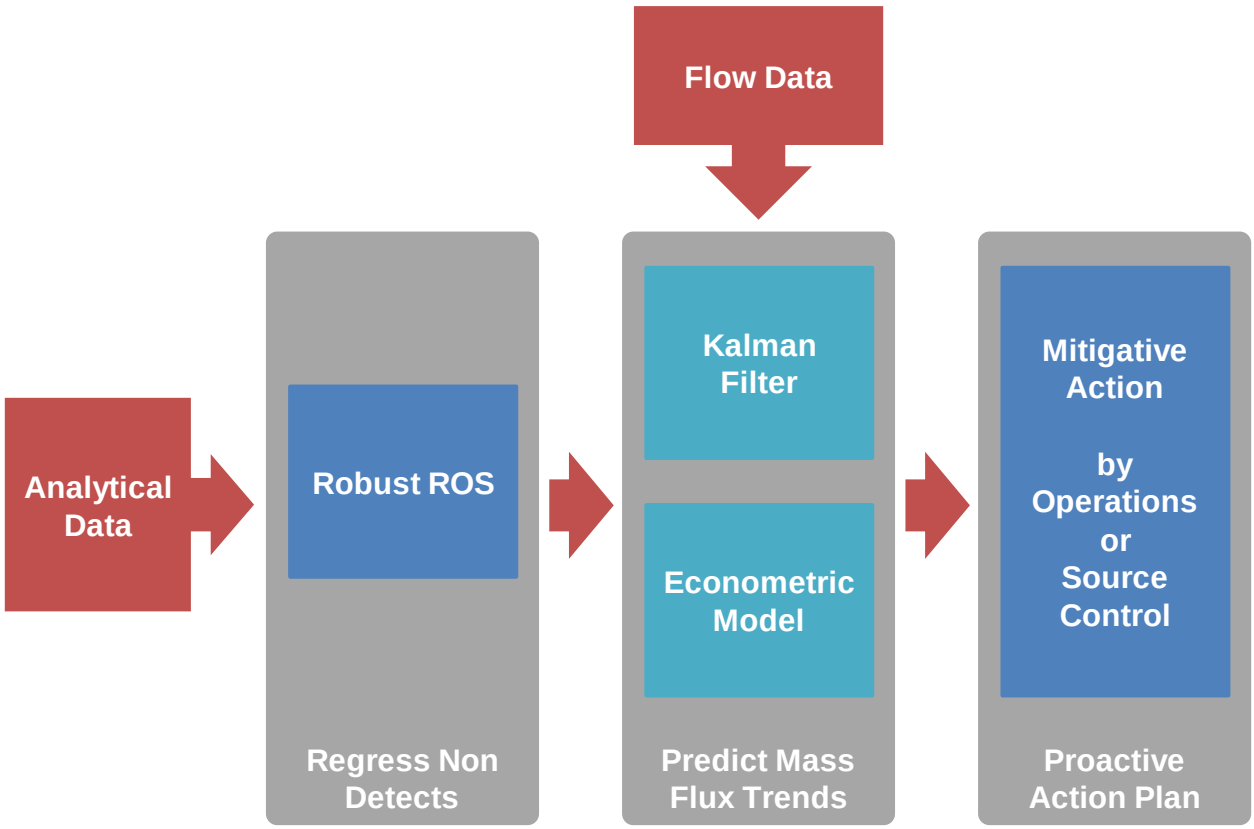

Figure 9. Predictive model for microconstituent concentration and source control.

ACKNOWLEDGMENTS

I wish to thank the following people. Without their support, this paper would not be possible: Jerry Evangelista, OCSD Source Control Supervisor; Mahin Talebi, OCSD Source Control Manager; Steve Fitzsimmons, OCWD Lab Director; and my wife, Lauren Viscardi.

REFERENCES

Bechmann, Henrik, et. Al. (1998). Control of sewer systems and wastewater treatment plants using pollutant concentration profiles. Water Science and Technology, 37(12), 87–93.

Bechmann, Henrik, et. Al. (1999). *Grey-box Modeling of Pollutant Loads from a Sewer System*, UrbanWater, 1(1), 71–78.

Bechmann, Henrik (1999). Modeling of Wastewater Systems, ATV Erhvervsforskerprojekt EF 623 IMM-PHD-1999-69, Technical University of Denmark, and Krüger A/S.

Bucy, R.S. and Joseph, P.D., Filtering for Stochastic Processes with Applications to Guidance, John Wiley & Sons, 1968; 2nd Edition, AMS Chelsea Publ., 2005.

Chui, Charles K. and Chen, Guanrong (2009). Kalman Filtering: with Real-Time Applications, Springer-Verlag, 4th ed., Belin Heidelburg, Germany, 229p.

Conerly, Octavia (2008). *Approaches to Screening for Risk From Pharmaceuticals in Drinking Water and Prioritization for Further Evaluation*, C-07-021, WA-B-02, Task 6, U.S. Environmental Protection Agency: Washington, D.C.

Dearden, John C., et. Al. (1989). The Development and Validation of Expert Systems for Predicting Toxicity, The Report and Recommendations of an ECVAM/ECB Workshop (ECVAM Workshop 24), ECVAM, TP 580, JRC Environment Institute, 21020 Ispra (VA), Italy.

Grewal, Mohinder S. and Andrews, Angus P. (2008). Kalman Filtering: Theory and Practice Using MATLAB, John Wiley and Sons; 3rd ed., Hoboken, New Jersey, 592p.

Hajek, Bruce (2009). An Exploration of Random Processes for Engineers, Notes for ECE 534, University of Illinois at Urbana-Champaign, Champaign, Illinois, 373p.

Helsel, Dennis R. (2005a). Nondetects and Data Analysis: Statistics for Censored Environmental Data. John Wiley and Sons, New York, 250p.

Helsel, Dennis R. (2006b). *Fabricating Data: How substituting values for nondetects can ruin results, and what can be done about it*. Chemosphere 65 (11), 2434-2439.

Helsel, Dennis R. (2005c). *Insider Censoring: distortion of data with nondetects*. Human and Ecological Risk Assessment 11, pp. 1127-1137.

Helsel, Dennis R. (2005d). *More Than Obvious: Better methods for interpreting nondetect data*. Environmental Science and Technol. 39 (20), p. 419A–423A.

Houseman, E. Andres (2004). A Robust Regression Model for a First-Order Autoregressive Time Series with Unequal Spacing: Technical Report, Harvard School of Public Health, ahousema@hsph.harvard.edu, 18p. Retrieved on 2008-04-23.

Kalman, R.E. (1960). "A new approach to linear filtering and prediction problems". Journal of Basic Engineering 82 (1): 35–45. Retrieved on 2008-05-03.

Shumway, R.H., Azari, R.S., and Kayhanian, M., (2002). *Statistical Approaches to Estimating Mean Water Quality Concentrations With Detection Limits*, Environmental Science and Technology 36, 3345-3353.

Singh, Anita and Nocerino, John (2006). Robust Estimation of Mean and Variance Using Environmental Data Sets with Below Detection Limit Observations,

Sorenson, H.W. (1970). Least Squares Estimation: From Gauss to Kalman, IEEE Spectrum, vol. 7, July, 1970, pp. 63-68.

Sorenson, H. W. (1980). Parameter Estimation, Principles and Problems, Mercel Dekker, New York, pp. 56-60.

Stacklin, Christopher (2008). *Pollutant Prioritization Project for Water Reuse*, 81st Annual

Water Environment Federation Technical Exhibition and Conference (WEFTEC 2008), Session 15, McCormick Place, Chicago, Illinois, October 20, 2008.

Stacklin, Christopher and Evangelista, Jerry (2008). *Technical Risk Assessment of Water Reuse in Consideration of Emerging Pollutants*, presented at the 2008 International Water Conference, Crowne Plaza - San Antonio Riverwalk, San Antonio, Texas, paper no. IWC-08-69, October 29, 2008.

Welch, Greg and Bishop, Gary (2006). An Introduction to the Kalman Filter, TR 95-041, Department of Computer Science, UNC-Chapel Hill, TR 95-041, July 24, 2006, pp. 1-16.

Whitehead, P. G. (Unknown). Water Quality Models for Waste Water Management, Australian National University, Centre for Resource and Environmental Studies, Canberra, Australia. 421-430

Winkler, Peter (2000). Optimization Heuristics in Econometrics: Applications of Threshold Accepting, John Wiley and Sons, West Sussex, England, 1st ed., 320p.

Zarchan, Paul and Musoff, Howard (2005). Fundamentals of Kalman Filtering: A Practical Approach (Progress in Astronautics and Aeronautics), American Institute of Aeronautics & Astronomy; 2nd ed., 764p.