EPA Public Access

Author manuscript

IEEE Syst J. Author manuscript; available in PMC 2019 June 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

IEEE Syst J. 2018 June ; 12(2): 1358–1368. doi:10.1109/JSYST.2016.2538082.

Developing a Model-based Drinking Water Decision Support System Featuring Remote Sensing and Fast Learning Techniques

Sanaz Imen,

Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL 32816 USA

Ni-Bin Chang,

Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL 32816 USA

Y. Jeffery Yang, and

U.S. EPA, Office of Research and Development, Water Supply and Water Resources Division, Cincinnati, OH, USA

Arash Golchubian

Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431

Abstract

Timely adjustment of operating strategies in drinking water treatment in response to water quality variations of both natural and anthropogenic causes is a grand technical challenge. One essential approach is to develop and apply integrated sensing, monitoring, and modeling technologies to provide early warning messages to plant operators. This paper presents a thorough literature review of the technical methods, followed by the development of a model-based decision support system (DSS). The DSS aims to aid water treatment operation via source water impact analysis. This model-based DSS featuring remote sensing and fast learning techniques can be easily applied by end-users and provide a visual depiction of spatiotemporal variation in water quality parameters of interest in source water. The system is able to forecast the trend of water quality one day into the future at a specific location and to nowcast water quality at water intake, thus helping the assessment of water quality in finished water against treatment objectives. The model-based DSS was assessed in a case study at a water treatment plant in Las Vegas, United States.

Index Terms-

Drinking Water; Machine Learning; Data Fusion; Remote Sensing; Forecasting Models; Decision Support Systems

I. INTRODUCTION

ACCESS to adequate and high quality water supply is one of the main requirements for a healthy life. Point-source and diffuse pollution degrade drinking water supply. Hence, drinking water treatment to remove contaminants is one of the greatest 20th achievements in public health protection. There are multiple treatment objectives, all depending on source water quality and timely adjustment of treatment operations. For example, higher levels of turbidity in the source water body is often associated with the presence of high levels of pathogens, for which coagulation-sedimentation-filtration are used along with chlorine for disinfection. However, chlorine-based disinfection also results in the formation of harmful disinfection byproducts when the level of dissolved organic carbon compounds is high. Therefore, real-time monitoring and projection of water quality are essential in treatment operations.

Water borne diseases from drinking water have long been recognized. Cholera and typhoid were one of the great 19th century threats facing the European and North American cities [1]. Recently, water pollution from microcystins (or cyanoginsins), a class of toxins produced by certain freshwater cyanobacteria, significantly challenged water treatment and supply in the western portion of Lake Erie, United States. The cyanobacteria bloom in August 2014 led to the shutdown of the water treatment plant and left more than 400,000 people in the city of Toledo without drinking water supply.

To detect contamination incidents in time and mitigate potential public health and economic consequences, the concept of a contamination warning system (CWS) was proposed and investigated in the mid-2000s [2]–[4]. The aim is to provide actionable information to drinking water treatment and distribution system operators. This type of CWS often relies on online sensors, rapid communication technologies, and data analysis methods [5]. During the last few years, significant progress has been achieved with the aid of smart sensors for monitoring pollutants to protect source water and improve treatment effectiveness. In general, smart sensors vary in functionality and applications. Some examples are biosensors and coagulation sensors for sewage treatment plants [6], and collars with a global positioning system (i.e., a bio-logging device) for cattle monitoring in pastures and real-time information retrieval of water quality conditions via sensor networks in lakes, rivers, and estuaries [7]. It is noted that more types of smart sensors have been used for water quality monitoring in expanded applications. Reported examples include the self-powered mobile sensors for water quality monitoring at drinking water distribution networks [8], a neural network-based software sensor for coagulation control [9], and a self-organized wireless biomedical microelectro-mechanical sensor system for autonomous control [10].

In addition to the advances of sensing technology, several optimization models were also developed to optimize sensor deployment in large-scale water distribution systems [11]–[16]. In 2011, a rule-based expert system was developed and compared to a few optimization models for sensor deployment in a small drinking water network [17]. A rule-based decision support system was then developed to analyze and generate a set of sensor placement locations and compare the performance against 10 optimization models [18]. Although these technologies are able to detect water contamination online and in real time for part of the

“drinking water infrastructure system”, they only cover a specific contamination event [19]. More research and development is needed to integrate sensors as an interactive water quality monitoring system in decision support, and to enable a system-wide monitoring domain from watershed protection to water treatment plant and then to drinking water distribution systems (Fig.1).

Large-scale areal remote sensing becomes feasible because of advances in satellites, communication networks, and sensor technologies. Satellite sensors with improved spatial, temporal, and spectral resolution allow broad applications in assessing and monitoring a suite of water quality parameters through machine learning algorithms (e.g., artificial neural network, genetic programming, support vector machine, etc.) [20]–[22]. Water quality forecasting allows prediction of pollution levels and enables action plans in advance. Developed water quality forecasting models vary from methods that handle linear (e.g., autoregressive integrated moving average (ARIMA)) to nonlinear (e.g., artificial neural network (ANN)) time series data [23]–[27]. Models can also be integrated as hybrid water quality forecasting models simultaneously handling both time series data with linearity and nonlinearity [28]–[30]. These integrated modeling schemes were mostly applied to predict water quality in a near real-time basis and/or to fill in missing data when regular monitoring systems were not functional in support of decision making. However, there is still a challenge to develop forecasting models to predict water quality a day ahead, allowing time to adjust drinking water treatment operation timely in response to contamination detected in the source water. The desired capability of an early warning system (EWS) is thus defined as “an integrated system for monitoring, interpreting, and communicating data, which can then be used to make decisions that protect public health and minimize unnecessary concern and inconvenience to the public” [4]. Such EWSs must be robust and sensitive to water quality changes of significance, be verifiable and affordable in cost, yield minimal false positive or negative responses, require low skill and training, provide a rapid response, and allow remote operation [31]. In the context of water supply, an EWS could be designed to monitor drinking water intakes and all inflows into the source water for water quality change detections. These sensors can also be networked with sensors in water treatment plants and drinking water distribution networks for cohesive interactions. Based on this requirement, a model-based decision support system (DSS) can be designed to provide water treatment plant operators with pollution alerts that would allow them to implement emergency response on site [32] or adjust operations to meet predefined treatment objectives. To this end, this study developed a water quality forecasting method capable of predicting multiple water quality parameters in source water one day in the future.

Advanced sensing technologies, such as satellite remote sensing, can enhance the setup of the DSS through continuous water quality monitoring with high spectral, spatial, and temporal resolution images [32]. Local data collection through the use of a surface water quality monitoring network may serve as ground truth for conducting remote-sensing-based water quality forecasting analyses and decision support, leading to alternative scenarios in drinking water infrastructure assessment [33].

A DSS may provide an interactive capability to assess potential drinking water consequences by processing data, integrating models, analyzing information, retrieving

knowledge, and visualizing results as a whole to better inform the decision makers than can data alone [33], [46]. Such a DSS may be organized as a comprehensive water quality management program linking upstream impact of potential source water quality changes to downstream management in water treatment and distribution networks.

Investigation of computerized DSSs began 50 years ago. First, the model-driven DSS was built in the late 1960s, and some theoretical developments then were performed in the 1970s [47]. Spreadsheet-based DSSs and Group DSSs, which were a new category of software to support group decision-making, were developed in the early and mid-1980s [27]. The addition of data warehouses, executive information systems, and business intelligence evolved in the late 1980s and early 1990s. Finally, the knowledge-driven DSS and the implementation of Web-based DSSs began in the mid-1990s [47]. Despite the numerous DSSs developed for water quality management (Table I), there is a lack of site-specific DSS that utilizes local sensors and satellite imagery data to estimate water quality parameters at a drinking water intake and assess possible removal efficiency in the water treatment process.

In this paper, the developed DSS is described with a focus on source water quality monitoring. It is then followed by an illustration of DSS architecture and an application in a case study of Lake Mead, located near the large metropolis of Las Vegas, Nevada, U.S.

II. Design of the Proposed DSS

The proposed DSS in the current study establishes relationships between surface spectral reflectance values of remote sensing and limited point concentration measurements of water quality parameters to estimate the water quality conditions throughout the lake. The technique relies on data fusion and inverse modeling for daily high-resolution monitoring. Satellite images collected by Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat satellite sensors were used for data fusion [48]. The end users of this DSS will be water treatment plant operators with a general familiarity of water quality management principles. The power of the developed DSS is its ability to provide a model-based decision support with the aid of remote sensing images and fast learning models for decision making. Visualization design makes the system user-friendly and easy to understand using the map-based display of results. The DSS provides end users with decision analysis to respond to events in order to achieve drinking water quality in compliance with the U.S. drinking water standards. Various software applied in the developed DSS are shown in Table II.

III. The Proposed DSS Architecture

The main objectives of the proposed DSS are to provide near real-time monitoring to examine the spatial variability of the desired water quality parameters within a lake, and forecast the concentrations of the selected water quality parameter at any specific location within a lake one day in the future. This data information is the necessary basis to evaluate mitigation measures through adjustment of treatment plant operations. For this purpose, the proposed DSS has three main modules: (1) graphic user interface, (2) data management, and (3) model management (Fig. 2).

The data management module is responsible for updating data and processing satellite imagery via two sub-modules, “Image processing” and “Database”. The model management module is empowered by a data fusion capability that merges two or more satellite imageries to improve spatial and temporal resolutions. The model management module provides the end user with formulation of modeling analyses, interprets results via the graphic user interface, and enable water treatment plant modeling with projected source water qualities. To achieve these tasks, the model management module is equipped with three sub-modules: the “Nowcasting Model”, “Forecasting Model”, and “Decision Analysis”. A detailed description of each module is presented in the following sections. The types of transferred data between different modules are shown in Fig. 2. Arrows in this figure indicate interaction between different modules and sub-modules (Fig 2).

A. Data Management Module

The data management module is responsible for updating data and processing satellite imagery. The Database sub-module is designed to reserve and update data received from the local water quality monitoring devices [e.g., a total organic carbon (TOC) analyzer] and user inputs.

Six types of data are reserved in the database: (1) processed satellite imagery, (2) historical fused images, (3) historical concentration maps, (4) concentration of water quality parameters, (5) historical reservoir elevations, and (6) unprocessed satellite images (Fig. 3).

All types of data except satellite imagery are used as inputs into the model management module (Fig. 3). It should be noted that the produced concentration maps and fused images are added to a database after completion of the nowcasting and/or forecasting modeling analyses. The second sub-module of the data management module which is titled Image Processing is capable of re-projecting, resampling, and cropping satellite images using ArcGIS. In addition, this sub-module can check the size of processed images and visualize red-green-blue images to check the cloud coverage of the site for the selected date. The ability to check and select the cloud-free images is offered to the user in this module as a requirement of the model management module.

B. Model Management Module

The model management module offers the end user an interface to formulate the models, analyze the results, interpret the modeling outputs, inform the user of important water quality parameters with regulatory implications, and propose action plans. To achieve these goals, the model management module includes three sub-modules. Whereas the nowcasting model may generate concentration maps to provide the user with detailed information about the spatiotemporal variation of the selected water quality parameters, the forecasting model may forecast the selected water quality parameter at the specific location within a lake one day in the future. Decision analysis is also applied to respond to the need for an automatic reaction (e.g., alarm triggering) to certain conditions or events (e.g., sudden changes in influent concentrations). The structures of the aforementioned sub-modules and their applied methodology are described in detail in the following sections.

Nowcasting Model—This sub-module starts with the user's selection of a date and type of water quality parameter (Fig. 4). First, the availability of a concentration map for the selected date and parameter in the database is checked, and if available, the user can view the map. Otherwise, the fused image or Landsat image alone for the selected date is assessed. If a fused image or Landsat image is available for the selected date, the concentration map is generated using the spectral reflectance value of each pixel of fused image or Landsat image as inputs into the ANN model for inverse modeling.

An ANN model developed by this method should have been well trained for the selected water quality parameter. This procedure is repeated until all pixels within a lake are covered. Once the procedure is completed, the estimated concentration values of the selected water quality parameter are organized as a concentration map. Then the map is added to the database and displayed to the user.

If the fused image or Landsat image of the selected date is not available in the database, then two pairs of the processed MODIS and Landsat images taken before and after the date defined by the user are loaded from the database. Although the daily revisit time of MODIS provides high temporal resolution, land bands of MODIS have poor spatial resolution. However, Landsat TM/ETM+, with a 16-day revisit cycle, has a high spatial resolution of 30 m. Therefore, the developed DSS is equipped with the spatial and temporal adaptive reflectance fusion model (STAR-FM) to generate a fused image with both high spatial and temporal resolutions whenever necessary [48].

Following these steps, the fused image is saved to the database for later use. The concentration map for the selected water quality parameter and date is then produced using the spectral reflectance value of each pixel of the produced fused image as inputs into the trained ANN model to estimate the concentration value of that pixel in support of nowcasting. The generated concentration map is added into the database and visualized by the end user to explore the probable hot spots within a lake or surface water body. This information may help watershed managers identify pollutant sources and recommend management strategies. These management strategies are defined by knowledge derived from (1) review articles, theses/dissertations, and reports related to the detected pollutant sources in the study area; (2) analysis of available historical water quality parameters of a lake; and (3) meetings and knowledge obtained from experts from a drinking water treatment plant. A sample of a list of the proposed management strategies is shown in Table III.

Forecasting Model—Two types of models were applied in the forecasting sub-module. They include the satellite remote-sensing-based and the statistics-based forecasting model. The former requires different types of data (i.e., selected water quality parameter, remote sensing reflectance values, and lake/reservoir elevations) capable of performing water quality forecasting for the entire lake. Although these multi-faceted input parameters allow the satellite remote-sensing-based forecasting model to predict desired water quality parameters with reasonable accuracy, it takes time to process the remote sensing data before they are available to the forecasting model. To bridge this gap, the latter was developed and employed within this sub-module. In case there is a need to carry out emergency response on

cloudy days when there is no clear remote sensing image, the second forecasting model only requires using the historical record of the desired water quality parameter via local water quality sensors to generate forecasting based on quantile information in statistics. This sub-module is able to provide the end users with the forecasted values of the desired water quality parameter in a very short time.

In the graphic user interface of this DSS, the satellite remote-sensing-based forecasting model started with selecting the desired water quality parameter and clicking on the forecasting tab. First of all, all available fused images in the last 30 days are loaded from the database, and a list of available images is displayed for the end users. If all required images are available, the forecasting procedure begins. Otherwise, the end user must select the image pairs for data fusion for the last 30 days to fill in the gaps of fused images in the database.

When the MODIS and Landsat images are available in the database for the selected pairs, images are used as inputs into STAR-FM to generate fused images for the selected period. Once all fused images are available for the selected time period, they are added to the database. Next, the surface reflectance values of all bands are extracted from the produced fused images at the location of the drinking water intake and used as inputs into the first forecasting model to estimate the concentration values of the selected water quality parameter. The estimated time-series concentration values of the selected water quality parameters are added to the database.

In the subsequent satellite remote-sensing-based forecasting, the updated daily time series of concentration values for the selected water quality parameter, reflectance values, and reservoir/lake elevations of the last 30 days are used as inputs into the nonlinear autoregressive neural network with external input (NARXNET) model to predict the concentration value for the next day (Fig. 5 and Eq. 1). The output of NARXNET depends on not only the time series of input variables but also the outcome of the model that may in turn become the input of the model iteratively to improve the learning capacity. This fast learning equation can be written as follows [49]:

$$y(t) = f(y(t-1), \dots, y(t-d), x(t-1), \dots, x(t-d)) \quad (1)$$

where $y(t)$ is the predicted value of the selected water quality parameter at time t ; $x(t)$ is the time series for each of the input variables (i.e., selected water quality parameter, remote sensing spectral reflectance values and reservoir elevations); and d is the feedback delay at the input and node. Note that the end users are supposed to update the database with the current value of reservoir elevation to address the drought or flood impact as inputs to support NARXNET forecasting analysis.

The statistics-based forecasting model is aimed to forecast the quantile one day into the future while considering the recent past probability distribution of the desired water quality parameter. The time series of the selected water quality parameter, which is used to calculate the quantiles, can be obtained from the archived ground-truth sampling record and/or the *in*

situ measurements via existing online sensors. Note that for the dates without sampling, the remote-sensing-based data fusion process, as mentioned in the previous forecasting model, can help expand the size of the database for probability analysis. It should be noted that the size of the selected water quality parameter is expanded to provide the most recent past days (i.e., the last 30 days). Given the set of values of the selected water quality parameter $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$, the quantile is calculated. The first step to define the quantile is to sort the original samples into statistical order as follows:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} \quad (2)$$

Then, the quantile corresponding to each value in the statistical order can then be generated based on the following equation:

$$P_i = \frac{i-1}{n-1} \quad i = 1, 2, \dots, n \quad (3)$$

The calculated quantiles of the last 30 days are used as input into an ANN model to predict the quantile for the next day. The structure of the ANN model is aimed to forecast the target water quality variable as future values of the input time series (i.e., quantiles of the recent last days). The true value of the selected water quality parameter corresponding to the predicted quantile needs to be converted after predicting the quantile for the next day. For this purpose, linear interpolation between the two nearest quantiles is applied to obtain the true water quality value:

$$x_k = \frac{(Q(P_k) - Q(P_i))}{(Q(P_{i+1}) - Q(P_i))} (x_{i+1} - x_i) + x_i \quad (4)$$

where x_i and x_{i+1} are the values of water quality parameters corresponding to $Q(P_i)$ and $Q(P_{i+1})$, respectively; $Q(p_i)$ and $Q(P_{i+1})$ are the first quantiles greater and less than the predicted quantile, respectively; x_k is the predicted value of water quality parameter; and $Q(P_k)$ is the predicted value of quantile.

Decision Analysis—Once the desired water quality parameters at particular locations have been estimated, they are sent to the decision analysis sub-module for possible optimization of the water treatment process via the use of the water treatment plant climate change adaptation model (WTP-ccam) (Fig. 6).

WTP-ccam is a simulation model package to simulate water quality for each water treatment unit operation or the entire treatment train. Built upon the original water treatment plant model [56], WTP-ccam incorporates a Monte Carlo engine to quantify probability variations in source water quality based on a set of new mechanistic process models for water quality analysis. For example, a calibrated logistic model for granular activated carbon (GAC) unit process is applied to simulate the changes of the finished water quality parameter given the

estimated or assumed range of water quality parameters at the drinking water intake. Model parameters are estimated based on a nonlinear regression algorithm [57]. With the aid of the calibrated logistic model, the employed GAC units in a water treatment plant can be operated in such a way that mitigates the effects of sudden changes in influent concentrations of TOC and turbidity. Unit process can be changed to produce finished water meeting the U.S. EPA drinking water standards. As shown in Fig. 6, the application of this model will aid the treatment plant operators in making better decisions via integrated sensing, networking, and control in the DSS.

IV. Risk Management

Better understanding of the potential causes of pollution occurrence and possible contaminant distribution in surface water body helps generate risk management strategies to improve drinking water safety (Fig. 7). Two core components may be considered at the impact identification stage: (1) models to estimate concentration values in near real time or by one day ahead; and (2) maps to indicate water quality variations in space and time or identify the impact areas of pollution events. After carrying out the impacts identification, the potential impacts to water supply are assessed in the “Impact Assessment” step that requires knowledge of the types of impacts and the time period of the occurrence. Types of impacts are categorized into (1) direct impacts, (2) indirect/secondary impacts, and (3) cumulative impacts [50]. The secondary impacts may be the effects of direct impacts occurring at other hydrologically connected locations, while cumulative impacts occur over time and space from number of developments and activities [50]. In addition, the impact duration can be categorized as short, medium, long-term, or permanent [50]. After assessing the impacts, impact management is performed to include risk prevention and mitigation strategies by (1) managing water pollution at the source by minimizing the contaminant loads into the source water, and (2) enhance contaminant removal strategies at the water treatment plant.

V. CASE STUDY OF THE PROPOSED DSS

Practical implementation of the new DSS was assessed by a case study at Lake Mead, the largest man-made reservoir in the southwestern U.S. The lake formed by the Hoover Dam along the Colorado River (Fig. 8) is the main source water for municipalities at the border of Nevada and Arizona, U.S.

A. Description of the Water Quality Condition

There are three major inflows into the lake: the Colorado River in the east, the Muddy and Virgin rivers in the north, and the Las Vegas Wash in the west (Fig. 8). Forest fire events on the northern side of the lake [48], [51] potentially increase total suspended solids (TSS) and TOC concentrations. In addition, urban stormwater runoff and treated wastewater effluent of different water quality flow into the lake through the Las Vegas Wash. Colorado River transports a large amount of sediments into Lake Mead, although the sediment inflow has decreased by 90% since the completion of Glen Canyon Dam in 1963, which is located upstream of Hoover dam. In addition, the water level of Lake Mead dropped 33 m from

2000 to 2008 as a result of the long-term drought. Because of water stratification in the lake, the decrease in water elevation or lake volume has resulted in large changes in water quality.

TOC in the water phase reacts with chlorine-based disinfectants to form disinfection by-products (DBPs) such as Trihalomethanes (THMs) in water treatment processes. TOC monitoring in the lake and around the water intake can help adjust treatment processes to minimize the DBP formation. For this purpose, the new DSS was used to examine TOC and TSS concentrations within the Lake Mead and to generate potential treatment options.

B. Insertion of the Data into the Database

Two types of satellite imagery (i.e., MODIS and Landsat TM/ETM+) and two water quality parameters (i.e., TOC and TSS) were selected for demonstration. Water quality parameters were obtained from Southern Nevada Water Authority (SNWA) for the time period of 2000–2013. TSS and TOC data were collected from 10 (i.e., 85 samples) and 26 water quality monitoring stations (i.e., 1252 samples), respectively. Landsat TM/ETM+ and MODIS (MOD09GA) land surface reflectance were obtained from the United States Geological Survey (USGS) and USGS MODIS Reprojection Tool Web Interface (MRTweb) for the time period of 2000–2013, respectively.

In this new DSS, the “Water Quality Parameters” tab provides the end users with a list of available sampling data for both TOC and TSS as well as the location of all available sampling stations within the lake (Fig. 9). The cursor can be moved over the lake to show the station name and number of available sampling data for that particular station. By clicking on each station (Fig. 9), the historical time series of TOC and TSS for the selected station can be observed. For the days with no sampling data, the values of TOC and TSS concentrations were left blank (Fig. 9). Stations are shown in different colors to distinguish stations that have TOC or TSS data. On the “Satellite Imagery” tab, the end users can observe a list of all available satellite images in the database and add more satellite images to the database from the repository folder in which the downloaded images were saved.

The database can be updated by the end users periodically (Fig. 3). They are able to not only input the recent sampling data collected from the available stations into the database, but also add a new sampling station into the database via the “locations” tab. To update the sampling data, the end users only need to click on the name of the sampling locations on the map and then add the new sampling data, and its corresponding date at the end of the time series shown in a tabular format (Fig. 9).

C. Satellite Imagery Processing

Processing satellite imagery is required prior to modeling analysis. The end users can check the availability of processed satellite imagery for a specific date via the Image Processing module and can add a processed image to the database by clicking on the process button, which automatically loads the downloaded images from the repository folder and runs Python scripts to process the images using ArcGIS. Once the satellite image processing is completed, the true color image is demonstrated on the screen to allow the user to check the cloud coverage of the selected image. This step is required for modeling because only cloud-free images can be applied for the modeling analysis.

D. Event-Driven Capability of the DSS

This DSS is event-driven enabling the end users to model water quality variations within the lake. The TOC and TSS concentrations may change within the lake, when the lake is influenced by exogenous events such as forest fire events in the watersheds, transported sediments from the Upper Colorado River, and the discharge of urban stormwater and treated wastewater effluents from Las Vegas Wash. The lake source water is treated at two treatment facilities using conventional flocculation, coagulation and filtration processes. High TOC and TSS concentration in the source water can affect operation of the conventional treatment units, leading to water quality variations and impacts on downstream water disinfection process.

This section is thus focused on applying the new DSS to assess the impacts of forest fire events on TSS concentration in the Overton Arm (Fig. 8) during one of the major forest fire events in 2005 [48]. The largest forest fire events occurred in April, May, June, and July 2005 in the upstream watersheds of Lake Mead [53]. The nowcasting model can run for each day of these months in which cloud-free satellite images are available. In addition to these months, this model can run for days before and after the forest fire events. Because the negative impacts of forest fire events on soil erosion may have a long lag time to manifest, assessing their impacts over the long-term is crucial.

A TSS concentration map in June 2005 shows high levels within the Overton Arm (Fig. 10). In this period, three monitoring stations CRLM, VR25.1, and LVB2.7, at the location of inflows from the Colorado River, Lower Virgin River Watershed, and Las Vegas Wash, respectively recorded different TSS concentrations (Fig. 8). The TSS level is categorized as High at the location of VR25.1, in contrast to the Normal levels at location of CRLM and LVB2.7 (Fig 10). The map also provides watershed managers with the data to develop management strategies in the watersheds. The high TSS concentration of the selected date in the Overton Arm is potentially related to the forest fire in the upstream watershed. Hence, improved watershed management to control the overland runoff and reduce soil transport becomes a potential management strategy (Fig. 10). Note that the current version of DSS does not provide the end users with detailed instruction of implementation for the proposed management strategies. The DSS can be further improved by including such information via meetings with the domain experts of water treatment and watershed management.

The DSS developed herein is also able to forecast TOC and TSS concentrations at the location of the drinking water intake. For Lake Mead, the recent past values of TOC and TSS concentrations were obtained for stations LVB6.7 and BB-7 near the drinking water intake (Fig. 8). In the forecasting sub-module, the end users are able to predict TOC or TSS values one day ahead by either using a remote-sensing-based forecasting model or a statistics-based forecasting model depending on the actual condition. As shown in Fig. 2, both forecasting and nowcasting sub-modules are connected to the decision analysis submodule where the WTP-ccam simulation can produce the corresponding strategies for essential adjustments of treatment processes as discussed above.

In this context, the reliability of the forecasting models was evaluated for selected time periods of past events (14 Aug–14 Oct 2001 and 11 Sep–28 Oct 2007). The evaluation is

based on the reproducibility of the DSS output compared to the available ground-truth data. For this purpose, the estimated time series of TOC in the selected time periods were applied to predict a TOC value one day ahead (15 Oct 2001 and 29 Oct 2007). In this statistics-based modeling analysis, the trained ANN model was applied to estimate the daily time series of TOC for the selected time period. The estimated and observed values for dates of ground-truth data are shown in Table IV for comparison. Advanced comparisons can be made possible across two types of forecasting models as listed in Table V. The two forecasting methods yield similar TOC projections, all of which are close to the actual observations.

VI. Final Remarks

In the near future, when more satellites such as Sentinel satellites are available, the fusion with different satellite imagery can be more powerful in remote sensing functionality, yielding stronger support for surface water quality monitoring and thus decision making in near real-time water plant operations. The upgraded DSS is applicable for large surface water bodies, such as western part of Lake Erie where microcystins blooms have impacted near-by water treatment plant operations [54], and for small drinking water reservoirs such as Harsha Lake, a source water for the McEwen Water Treatment Plant in Ohio [55].

VII. Conclusion

This paper provides a thorough literature review, and presents a new DSS for a drinking water treatment plant using surface water as the source. The new DSS featured by remote sensing and fast learning algorithm enables water treatment plant operators to visualize spatiotemporal variations of the water quality parameters in source water and to predict the selected water quality parameter at specific locations. In addition, the prototype of this model-based DSS has the capability of enabling the water treatment plant operator to simulate the finished water quality for the given source water quality variations, and hereby to develop corresponding adjustments for treatment processes. Practical implementation of the new DSS was assessed by a case study in the water treatment plant at Lake Mead, which provides drinking water for more than 25 million people in the Southwestern U.S. Results demonstrate the benefit of such a model-based DSS to the drinking water treatment plant via multi-sensor acquisition, image fusion, machine learning, and alarm triggering for sustainable decision making of drinking water infrastructure in a fast-growing urban region – Las Vegas. Such demonstration gives rise to insight about how integrated sensing, networking, and control can cohesively improve the sustainability in our urban life line systems.

Acknowledgment

The authors acknowledge the help from Ms. Susan Holmes in SNWA for providing us with Lower Colorado River water quality database and Mr. Mao Fang in Las Vegas Valley Water District for the helpful information of the water treatment plant. In addition, the authors acknowledge the help from Mr. Benjamin Vannah and Dr. Kaixu Bai for helping develop the MATLAB code. The manuscript is based on investigation results funded by the U.S. EPA through contract EP-C-11-006. It has been subjected to the administrative review by US EPA and has been approved for external publication. Any opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Agency; therefore, no official endorsement should be inferred. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Biographies

Sanaz Imen received her B.S. degree in civil engineering from the University of Science and Culture in Iran, in 2005, and M.S. in civil engineering-planning and management of water resources from the University of Tehran in Iran, in 2007. She received her Ph.D. in civil engineering-water resources engineering program from the University of Central Florida, in 2015. Her research interests include climate change effects on water quality management in river-reservoir system, water quality modeling, and environmental remote sensing.

Y. Jeffrey Yang received the B.Sc. degree in geomechanics from China University of Geosciences, Beijing, China, and the M.Sc. degree in geochemistry from Chinese Academy of Geological Sciences, Beijing, China, and the Ph.D. degree in isotope geochemistry from Miami University, Oxford, OH, USA. He is a Research Scientist with the Office of Research and Development, EPA, USA.

Ni-Bin Chang (SM'10) received the B.Sc. degree in civil engineering from the National Chiao-Tung University, Hsinchu, Taiwan, and the M.Sc. and Ph.D. degrees in environmental systems engineering from Cornell University, Ithaca, NY, USA, in 1983, 1989, and 1991, respectively. He is a Professor with the University of Central Florida, Orlando, FL, USA. His research interests include environmental systems optimization, integrated data fusion and mining for environmental, ecological, and hydrological applications.

Arash Golchubian received his B.Sc. degree in Computer Engineering from Florida Atlantic University. He works for Motorola Solutions Inc. for eight years and is a Senior Software Architect working on distributed Windows applications.

REFERENCES

- [1]. Fawell J and Nieuwenhuijsen MJ, "Contaminants in drinking water," Br. Med. Bull, 68, pp. 199–208, 2003. [PubMed: 14757718]
- [2]. ASCE American Society of Civil Engineers, "Interim voluntary guidelines for designing an online contaminant monitoring system," p.p. 557, 2004.
- [3]. Roberson JA and Morley KM, "Contamination warning system for water: an approach for providing actionable information to decision-makers," American W. Denver, CO, 2005.
- [4]. U.S. EPA, "Technologies and techniques for early warning system for water: an approach for providing actionable information to decision-makers: a state-of-the-art review," Office of Research and Development - National Homeland Security Research Center, Water Infrastructure Protection, 2005.
- [5]. U.S. EPA, "Sensor network design for drinking water contamination warning systems," Office of Research and Development - National Homeland Security Research Center, Water Infrastructure Protection, 2010.
- [6]. Umetsu J, "Instrumentation, control and sensor technology for sewage treatment plants," Fuji Electr. Rev, 45, no. 4, pp. 126–130, 1999.
- [7]. Oppermann I, "Smarter water management through sensor-based networks," Session 3A: Smart Technologies. [Online]. Available: <http://www.oecd.org/site/stitff/presentationsandwebcasts.htm>.
- [8]. Wu R, Wan Salim WWA, Malhotra S, Brovont A, Park JH, Pekarek SD, Banks MK, and Porterfield DM, "Self-powered mobile sensor for in-pipe potable water quality monitoring," in 17th International Conference on Miniaturized System for Chemistry and Life Science, Freiburg, Germany, 2013, pp. 14–16.
- [9]. Valentin N and Deneux T, "A neural network-based software sensor for coagulation control in a water treatment plant," Intell. Data Anal, 5, pp. 23–39, 2001.
- [10]. Beferull-Lozano B, "Autonomous control of large-scale water treatment plants based on self-organized wireless BioMEM sensor and actuator networks," [Online]. Available: http://cordis.europa.eu/project/rcn/99693_en.html.
- [11]. Berry J, Hart J, Phillips W, and Uber C, "A general integer-programming-based framework for sensor placement in municipal water networks," in Critical Transitions in Water and Environmental Resources Management, 2004, pp. 1–10.
- [12]. Dorini D, Jonkergouw G, Kapelan P, Di Pierro Z, Khu F, and Savic S, "An efficient algorithm for sensor placement in water distribution systems," in Water Distribution Systems Analysis Symposium, 2006, pp. 1–13.
- [13]. Ghimire B, and Barkdoll S, "Heuristic method for the battle of the water network sensors: demand based approach," in Water Distribution Systems Analysis Symposium, 2006, pp. 1–10.
- [14]. Guan W, Aral J, Maslia M, and Grayman M, "Optimization model and algorithms for design of water sensor placement in water distribution systems," in Water Distribution Systems Analysis Symposium, 2006, pp. 1–16.
- [15]. Krause P, Leskovec A, Isovitsch J, Xu S, Guestrin J, VanBriesen C, Small J, and Fischbeck M, "Optimizing sensor placements in water distribution systems using submodular function maximization," in Water Distribution Systems Analysis Symposium, 2006, pp. 1–17.
- [16]. Wu T, and Walski Z, "Multi-objective optimization of sensor placement in water distribution systems," in Water Distribution Systems Analysis Symposium, 2006, pp. 1–11.
- [17]. Chang NB, Pongsanone NP, and Ernest A, "Comparisons between a rule-based expert system and optimization models for sensor deployment in a small drinking water network," Expert Syst. Appl, 38, no. 8, pp. 10685–10695, 2011.
- [18]. Chang NB, Prapinpongsonone N, and Andrew E, "Optimal sensor deployment in a large-scale complex drinking water network: Comparisons between a rule-based decision support system and optimization models," Comput. Chem. Eng, 43, pp. 191–199, 2012.
- [19]. van Wijlen MAB, Koerkamp MK, Xie RJ, Puah AN, van Delft W, Bajema B, and Verhoef JW, "Innovative sensor technology for effective online water quality monitoring," in Proceedings of the 4th Singapore International Water Week, Singapore, 2011, pp. 1–33.

- [20]. Panda SS, "Artificial neural networks application in lake water quality estimation using satellite imagery," *J. Environ. Informatics*, 4, no. 2, pp. 65–74, 12 2004.
- [21]. Doña C, Chang N-B, Caselles V, Sánchez JM, Camacho A, Delegido J, and Vannah BW, "Integrated satellite data fusion and mining for monitoring lake water quality status of the Albufera de Valencia in Spain," *J. Environ. Manage*, 151, pp. 416–426, 2015. [PubMed: 25602695]
- [22]. Huang W, Huang F, and Song J, "An SVM model for Water Quality Monitoring Using Remote Sensing Image," in *Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10) Jinggangshan*, P. R. China, 2–4, 4 2010, pp. 186–189.
- [23]. Parmar KS and Bhardwaj R, "Water quality management using statistical analysis and time-series prediction model," *Appl. Water Sci*, 4, no. 4, pp. 425–434, 2014.
- [24]. Palani S, Liong SY, and Tkalic P, "An ANN application for water quality forecasting," *Mar. Pollut. Bull*, 56, no. 9, pp. 1586–1597, 2008. [PubMed: 18635240]
- [25]. Gazzaz NM, Yusoff MK, Aris AZ, Juahir H, and Ramli MF, "Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors," *Mar. Pollut. Bull*, 64, no. 11, pp. 2409–2420, 2012. [PubMed: 22925610]
- [26]. Khalil BM, Awadallah AG, Karaman H, "Application of artificial neural networks for the prediction of water quality variables in the Nile Delta," *J. Water Resour. Prot*, 2012, no. 4, pp. 388–394, 2012.
- [27]. Heydari M, Olyaie E, and Mohebzadeh H, "Development of a Neural Network Technique for Prediction of Water Quality Parameters in the Delaware River, Pennsylvania," *Middle-East J. Sci. Res*, 13, no. 10, pp. 1367–1376, 2013.
- [28]. Alizadeh MJ and Kavianpour MR, "Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean," *Mar. Pollut. Bull*, 98, no. 1–2, pp. 171–178, 2015. [PubMed: 26140748]
- [29]. Ömer Faruk D, "A hybrid neural network and ARIMA model for water quality time series prediction," *Eng. Appl. Artif. Intell*, 23, no. 4, pp. 586–594, 2010.
- [30]. Liu S, Xu L, Jiang Y, Li D, Chen Y, and Li Z, "A hybrid WA-CPSO-LSSVR model for dissolved oxygen content prediction in crab culture," *Eng. Appl. Artif. Intell*, 29, pp. 114–124, 2014.
- [31]. Brosnan TM, Ed., "Early warning monitoring to detect hazardous events in water supplies," ILSI, 1999.
- [32]. Quansah JE, Engel B, and Rochon GL, "Early warning systems : a review," *Terr. Obs*, 2, no. 2, pp. 24–44, 2010.
- [33]. Booth NL, Everman EJ, Kuo IL, Sprague L, and Murphy L, "A web-based decision support system for assessing regional water-quality conditions and management actions," *J. Am. Water Resour. Assoc*, 47, no. 5, pp. 1136–1150, 2011. [PubMed: 22457585]
- [34]. Câmara AS, Cardoso da Silva M, Carmona Rodrigues A, Remédio JM, Castro PP, Soares de Oliveira MJ, and Fernandes TF, "decision support system for estuarine waterquality management," *Journal of Water Resources Planning and Management*, 116, no. 3 pp. 417–432, 1990.
- [35]. Yakowitz D, Stone J, Lane LJ, Heilman P, Masterson J, Abolt J, and Imam B, "A decision support system for evaluation the effects of alternative farm management systems on water quality and economics," *Wat. Sci. Tech*, 28, no. 3–5, pp. 47–54, 1993.
- [36]. Srinivasan R and Engel, "A spatial decision support system for assessing agricultural nonpoint source pollution," *J. Am. water Resour. Assoc*, 30, no. 3, pp. 441–452, 1994.
- [37]. Srinivasan R and Engel BA, "A spatial decision support system for assessing agricultural nonpoint source pollution," *Water Resour. Bull*, 30, no. 3, pp. 441–452, 1994.
- [38]. Osmond DL, Cannon RW, Gale JA, Line DE, Knott CB, Phillips KA, Turner MH, Foster MA, Lehning DE, Coffey SW, and Spooner J, "WATERSHEDSS: A decision support system for watershed-scale nonpoint source water quality problems," *J. Am. Water Resour. Assoc*, 33, no. 2, pp. 327–341, 1997.
- [39]. Chen CW, Herr J, Ziemelis L, Goldstein RA, and Olmsted L, "Decision support system for total maximum daily load," *J. Environ. Eng*, 125, pp. 653–659, 1999.

- [40]. Leon LF, Lam DC, Swayne DA, Farquhar GJ, and Soulis ED, "Integration of a non-point source pollution model with a decision support system," *Environmental Model. Softw*, 15, pp. 249–255, 2000.
- [41]. Matthies M, Berlekamp J, Lautenbach S, Graf N, and Reimer S, "Decision support system for the Elbe River water quality management," *Proc. ModSim 2003*, Townsv., no. BfG, pp. 284–289, 2003.
- [42]. Fassio A, Giupponi C, Hiederer R, and Simota C, "A decision support tool for simulating the effects of alternative policies affecting water resources: an application at the European scale," *J. Hydrol*, 304, no. 1–4, pp. 462–476, 2005.
- [43]. Nasiri F, Maqsood I, Huang G, and Fuller N, "Water quality index: A fuzzy river-pollution decision support expert system," *J. Water Resour. Plan. Manag*, 133, no. April, pp. 95–105, 2007.
- [44]. Obropta CC, Niazi M, and Kardos JS, "Application of an environmental decision support system to a water quality trading program affected by surface water diversions," *Environ. Manage*, 42, no. 6, pp. 946–956, 2008. [PubMed: 18592303]
- [45]. Sharma A, Naidu M, and Sargaonkar A, "Development of computer automated decision support system for surface water quality assessment," *Comput. Geosci*, 51, pp. 129–134, 2012.
- [46]. B. NL Graffy EA, "Linking environmental risk assessment and communication: an experiment in co-evolving scientific and social knowledge," *Int. J. Glob. Environ. Issues*, 8, no. 1/2, pp. 132–146, 2008.
- [47]. Power DJ, "A brief history of decision support system," 2007 [Online]. Available: <http://dssresources.com/history/dsshstory.html>.
- [48]. Imen S, Chang N-B, and Yang YJ, "Developing the remote sensing-based early warning system for monitoring TSS concentrations in Lake Mead," *J. Environ. Manage*, 160, pp. 73–89, 2015. [PubMed: 26093101]
- [49]. Cai S, Kang X, and Zhang L, "Time series prediction based on artificial neural network for estimation of forest biomass," *Int. J. Adv. Comput. Technol*, 4, no. 12, pp. 318–326, 2012.
- [50]. P. M and R. T, Eds., "Methods of environmental impact assessment," Routledge, Taylor and Francis Group, 2009.
- [51]. Chen H, "Simulating the hydrologic impacts of land cover and climate changes under a semi-arid environment," University of Cincinnati, 2013.
- [52]. L. V. V. W. District, "Water quality report," 2013 [Online]. Available: www.lvvwd.com/assets/pdf/wqreport.pdf.
- [53]. N. I. C. S. (NICS), "2005 Statistics and summary- introduction and summary." [Online]. Available: http://www.predictiveservices.nifc.gov/intelligence/2005_statsumm/2005Stats&Summ.html.
- [54]. Chang NB, Vannah B, and Jeffrey YJ, "Comparative sensor fusion between hyperspectral and multispectral satellite sensors for monitoring microcystin distribution in Lake Erie," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*, 7, no. 6, pp. 2426–2442, 6 2014.
- [55]. Chang NB, Vannah BW, Yang YJ, and Elovitz M, "Integrated data fusion and mining techniques for monitoring total organic carbon concentrations in a lake," *Int. J. Remote Sens*, 35, no. 3, pp. 1064–1093, 2 2014.
- [56]. U.S. Environmental Protection Agency, "Water Treatment Plant Model Version 2.2 User's Manua.", Office of Ground Water and Drinking Water, U.S. Environmental Protection Agency, Cincinnati, Ohio, 2005.
- [57]. Li Z, Buchberger SG, Clark RH, Yang YJ, and Swertfeger J, "Evaluation of the logistic model for GAC performance in water treatment", *J. Am. Water Works Assoc*, 104, pp. E489–E500, 2012.

Fig. 1.
Role of the DSS in a water quality management system

Fig. 2.
Structure of the developed DSS

Fig. 3. Structure of the data management module; the blue box shows the data types used as inputs into the model management module

Fig. 4.
Structure of the nowcasting model

Fig. 5.
Structure of the forecasting model

Fig 6.
Structure of the decision analysis sub-module

Fig. 7.
The proposed strategy for the risk management

Fig. 8.
Location of Lake Mead and main river inflows into the lake [48]; CRLM stands for Colorado River/Lake Mead; VR stands for Virgin River; LVB stands for Las Vegas Bay

Fig. 9.
Data management module- database

Fig. 10.
The model management module – nowcasting

TABLE I.

Selected studies of applying decision support systems for water quality management

Reference	Geographic Focus	Objective
Câmara et al., 1990 [34]	Portugal	DSS named Hypertejo to interconnect databases and models to manage water quality of Tejo estuary
Yakowitz et al., 1993 [35]	Iowa, U.S.	Multiobjective decision support system (MODSS) to predict the impact of alternative management system on surface and groundwater quality as well as farm income
Srinivasan and Engel, 1994 [37]	Texas, U.S.	Special decision support system (SDSS) to assess agricultural nonpoint source pollution using nonpoint source pollution model and GIS
Osmond et al., 1997 [38]	-	Water, soil, and hydro-environmental DSS (WATERSHEDSS) to aid managers in defining water quality problems and selecting appropriate nonpoint source control measures.
Chen et al., 1999 [39]	North Carolina, and South Carolina, U.S.	DSS to calculate total maximum daily loads of various pollutants for water quality limited sections
Leon et al., 2000 [40]	-	Regional analysis information system (RAISON) DSS was linked with the agricultural nonpoint source pollution model (AGNPS) to deal with nonpoint source pollution modelling
Matthies et al., 2003 [41]	Germany	DSS for river basin management to calculate the long-term nutrient discharge, simulate the waste water pathways, and assess the aquatic fate
Fassio et al., 2005 [42]	Europe	Multi-sectoral, integrated, and operational decision support system (mDSS) to assess alternative measures for the reduction of nitrogen pressure from agriculture on water resources at the European level
Nasiri et al., 2007 [43]	Thailand	A fuzzy multiple-attribute DSS to compute the water quality index and provide an outline to prioritize the alternative plans.
Obropta et al., 2008 [44]	New Jersey, U.S.	Environmental DSS (EDSS) to address hot spots of concern for the water quality trading program, and select a water quality trading framework
Booth et al., 2011 [33]	South Atlantic Gulf and Tennessee River Basin, U.S.	Spatially referenced regression on watershed attributes (SPARROW) decision support system to display water quality conditions in a river basin and describe, test, and share modeled future scenarios
Sharma et al., 2012 [45]	India	Computer automated DSS-surface water quality assessment tool (SWQAT) to calculate overall index of pollution values and display it on Google map

TABLE II.

The List of required softwares for the developed DSS and their application

Software	Version	Application
Windows	7 and higher	Supported operating system
MATLAB	R2013a	To write image processing, nowcasting, and forecasting scripts
ArcGIS	10.2	Image processing
.NET framework	4.5	UI development
STAR-FM	1.1.2	Recompiled for Windows and used for producing fused images

TABLE III.

List of proposed management strategies to reduce the detected Contaminations

Pollutant Sources	Management Strategies
Grazing Livestock	Grazing management Riparian area management Nutrient management
Agricultural Runoff	Prevent plowing too often or at the wrong time Prevent improper, excessive, or poorly timed application of pesticides, irrigation water and fertilizer
Urban and Industrial Discharge	Properly treating wastewater
Erosion	Apply management practices to control the volume and flow rate of runoff water, reduce soil transport

TABLE IV.

Comparison between the estimated and observed TOC values at LVB6.7 USING the trained ANN model

Year	Date	Observed TOC (mg·L ⁻¹)	Estimated TOC (mg·L ⁻¹)
2001	28-Aug-2001	4.69	3.9
	4-Sep-2001	3.98	3.86
	25-Sep-2001	3.83	3.44
	7-Oct-2001	3.74	4.13
2007	11-Sep-2007	3.20	3.14
	17-Sep-2007	3.00	3.12
	25-Sep-2007	3.10	3.09
	1-Oct-2007	3.10	3.17
	8-Oct-2007	3.10	3.15
	15-Oct-2007	3.10	3.07
	22-Oct-2007	3.00	3.13

TABLE V.

Comparison between the results of forecasting models and the observed value of TOC at LVB6.7

Date	Observed TOC (mg·L ⁻¹)	Estimated TOC (mg·L ⁻¹)	
		Remote sensing-based	Statistics-based
15-Oct-2001	4.75	4.68	4.69
29-Oct-2007	3.11	3.23	3.14