Ecological Indicators 66 (2016) 428–439

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Improving real-time forecasting of water quality indicators with combination of process-based models and data assimilation technique

Xuan Wang, Jingjie Zhang*, Vladan Babovic

Department of Civil and Environmental Engineering, National University of Singapore, Singapore 119260, Singapore

ARTICLE INFO

Article history:
Received 24 October 2015
Received in revised form 29 January 2016
Accepted 2 February 2016

Keywords:
Water quality indicators
Modified local model
Water quality model
Data assimilation
Real-time forecasting

ABSTRACT

Water quality indicators can be used to characterize the status and quantify and qualify the change of aquatic ecosystems under different disturbance regimes. Although many studies have been done to develop and assess indicators and discuss interactions among them, few studies have focused on how to improve the predicted indicators and explore their variations in receiving water bodies. Accurate and effective predictions of ecological indictors are critical to better understand changes of water quality in aquatic ecosystems, especially for the real-time forecasting. Process-based water quality models can predict the spatiotemporal variations of the water quality indicators and provide useful information for policy-makers on sound management of water resources. Given their inherent constraints, however, such process models alone cannot actually guarantee perfect results since water quality models generally have a large number of parameters and involve many processes which are too complex to be efficiently calibrated. To overcome these limitations and explore a fast and efficient forecasting method for the change of water quality indictors, we proposed a new framework which combines the process-based models and data assimilation technique. Unlike most traditional approaches in which only the model parameters or initial conditions are updated or corrected and the models are run online, this framework allows the information extracted from observations and outputs of process models to be directly used in a data-driven local/modified local model. The results from the data-driven model are then assimilated into the original process model to further improve its forecasting ability. This approach can be efficiently run offline to directly correct and update the output of water quality models. We applied this framework in a real case study in Singapore. Two of the water quality indicators, namely salinity and oxygen were selected and tested against the observations, suggesting that a good performance of improving the model results and reducing computation time can be obtained. This approach is simple and efficient, especially suitable for real-time forecasting systems. Thus, it can enhance forecasting of water quality indictors and thereby facilitate the effective management of water resources.

1. Introduction

The worldwide deterioration of water quality due to human population growth, economic development and climate change is accelerated. Water quality indicators which reflect biological, chemical or physical attributes of ecological conditions can be used to characterize the status and quantify and qualify the change of aquatic ecosystems under different disturbance regimes (Jackson et al., 2000). Many studies have been done to develop and assess water quality indicators as well as discuss interactions among them (Jørgensen et al., 2010). For example, Cabecinha et al. (2009a) proposed to define the ecological status of reservoirs based on environmental variables and phytoplankton assemblages through multivariate analysis. Recatalá and Sacristán (2014) developed a minimum indicator set to evaluate natural resources quality at the municipality level. These indictors serve as a basis for assessing environmental impacts from land use planning instruments and from specific projects in the area of the European Mediterranean Region. However, few studies have focused on how to improve the predicted indicators and explore their variations in receiving water bodies. For example, Cabecinha et al. (2009b) extended Stochastic Dynamic Methodology in the management of the water quality to capture how expected changes at land use level will alter the water quality in a reservoir. Although such studies provide meaningful results on the interaction between ecological key-components and environmental conditions, it requires sufficient environmental

* Corresponding author at: Civil and Environmental Engineering Department, National University of Singapore, E1-08-25, No. 1 Engineering Drive 2, Singapore 117576, Singapore. Tel.: +65 82897619.
E-mail address: lakejz@gmail.com (J. Zhang).

http://dx.doi.org/10.1016/j.ecolind.2016.02.016
1470-160X/

datasets which normally remain difficult to obtain. An effective and accurate prediction of water quality indicators based on the available historical datasets and limited resources is critical to better understand changes of water quality in aquatic ecosystems, especially for real-time forecasting of water quality status and building-up an early-warning system.

To predict and analyze the status of water body, statistical or data-driven model is one popular tool, which has been widely used for many purposes, especially for exploring the relationship between different environmental indicators. For example, Pandey et al. (2012) used bivariate analysis to assess the linkage between *Escherichia coli* concentrations, watershed indexes and the rainfall condition. The multivariate regression models were applied to predict in-stream *E. coli* concentrations. However, such data-driven method is normally driven by the historical data set and does not take the internal processes of the indictors and variables into consideration. Therefore, its accuracy is sensitive and dependent on the amount and quality of available datasets. On the other hand, processed-based water quality model is an alternative which can provide useful information to understand the spatiotemporal variation of water quality. They have been used extensively to support planning and management of environmental systems (van Delden et al., 2011; Zhang and Jorgensen, 2005). However, given the constraints imposed by model resolution, parameter uncertainty, simplifying assumptions, lack of data etc., such process model alone cannot actually come without facing many challenges such as model development, integration and reliability (Gal et al., 2014). Efforts have been made to develop complex models for studying underlying processes in aquatic ecosystems (Liang et al., 2013). Moreover, aquatic ecosystems that normally involve many mechanisms such as physical, chemical and biological processes as well as numerous components are complex and impossible to be fully mimicked and covered in every respect (Zhang et al., 2003). The unconsidered processes and components in the model structure as well as the lack or missing of observation data may result in high uncertainties when models are used for predictions. Hence, calibrating models to observed outcomes is always difficult and requires an enormous amount of computation time (Zhang et al., 2003). Although super computers give a potential in enhancing the development of more and more complex process models, too complex model may not always guarantee better outcomes and does not appear necessary either for stakeholders when used for daily operation (Washington et al., 2009). A fast and efficient forecasting method for the change of water quality indictors is always a strong desire for policy-makers.

Given the advantages and limitations both for process-based water models and data-driven approaches, an combination of data assimilation with water quality model provides a feasible solution for better prediction of dynamics of water quality. Data assimilation is a methodology that can optimize the extraction of reliable information from observations and process-based models to improve the quality of estimation (Robinson et al., 1998). For a complex water quality model, it may be more convenient to apply a data assimilation scheme to correct the model without interfacing with and rerunning the water quality model. This efficiency is especially necessary for a real-time forecasting system.

Although various data assimilation techniques, including direct insertion method (DIR), Kalman filter (KF) and its derivative, have been used in environmental systems (Hartnack and Madsen, 2001; Haugen and Evensen, 2002; Reichle, 2008), most of them involve parameters or initial condition updating and correct the process model online when they are applied in and combined with process-based water quality models. Given a large number of parameters and many processes involved in the water quality model (as mentioned above), however, such online updating correction may be time-consuming and expensive and sometimes does not necessarily yield a better performance.

Therefore, in this paper we proposed an integrated modeling scheme which can directly update the output of the water quality model based on the combination of both water quality modeling and data-driven approach. This allows the model to be run offline and has advantages in terms of model simplification, efficiency and accuracy. Two methods based on chaos theory, i.e. local model (LM) and modified local model (MLM), were introduced and coupled with a water quality model to improve its forecasting capability. The scheme was tested on predicting two representative water quality indicators of salinity and dissolved oxygen. This kind of approach may offer one of the most comprehensive developments in the real-time forecasting of water quality indicators.

Thus, the objectives in this paper aim to: 1. introduce the new framework combining water quality model with data-driven model; 2. examine the performance of the proposed scheme with a real case study of Singapore Regional water body; 3. assess the influence of the observation reliability and process model on the performance of the proposed scheme; and 4. assess the impact of forecasting intervals on the predicted accuracy of the water quality model.

2. Methodology

2.1. Data-driven model

As mentioned before, in the integrated framework two data-driven methods, namely local model (LM) and modified local model (MLM), will be coupled with the process model to improve the forecasting capability. This approach is briefly described in the following.

2.1.1. Local linear model

Previous studies of the dynamic system revealed that many physical systems can be better interpreted using chaos theory (Williams, 1997). The random and irregular behavior in natural systems may arise from purely deterministic dynamics with unstable trajectories (Sun et al., 2009). A local linear model (LM) was developed based on the Taken's embedding theorem (Takens, 1981). This method can utilize the inner nonlinear deterministic rule in chaos system to reconstruct a phase (or embedded) space which is equivalent to original state from a scalar time series. It assumes that similar states will evolve similarly and an equal state has an equal future, and the configuration of the observed process can represent as a state in an appropriate phase through abstraction (Mancarella et al., 2008).

Given time series x_{t_n} at reference time point t_n, in order to properly present the underlying order of a dynamical system, a proper embedding should be created as $X_n(t) = \{x_{t_n}, x_{t_{n-\tau}}, \ldots, x_{t_{n-(d_e-1)\tau}}\}$. A Euclidean metric is imposed onto the phase space to find the k nearest neighbors, which are used to calculate the regression coefficient β_T, through which the scalar value can be finally forecast $\hat{x}_{t_n+\Delta t} = \beta_T \cdot X_n(t)$ (Refer to Babovic et al. (2005) for more details).

However, the predictive accuracy of the LM is sensitive to the initial condition and a slight deviation from a trajectory in the state space can lead to dramatic changes in future behavior (Guegan and Leroux, 2009). It hence causes reduction in the accuracy as the forecast horizon increases. Due to these limitations, a modified local model (MLM) is thus introduced as follows, which predicts the chaotic time series by invoking the chaos theory.

2.1.2. Modified local model

In order to mitigate the adverse influence caused by deviation from a trajectory in the initial state space, the MLM is proposed by Wang and Babovic (2014) by making use of a different chaotic