Ecological Informatics 28 (2015) 42–60

Contents lists available at ScienceDirect

Ecological Informatics

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

Diagnosis of the artificial intelligence-based predictions of flow regime in a constructed wetland for stormwater pollution control

Ni-Bin Chang [a,*], Golam Mohiuddin [a], A. James Crawford [a], Kaixu Bai [a,b], Kang-Ren Jin [c]

[a] *Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL, USA*
[b] *Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University, Shanghai 200241, China*
[c] *South Florida Water Management District (SFWMD), West Palm Beach, FL, USA*

ARTICLE INFO

Article history:
Received 14 February 2015
Received in revised form 4 May 2015
Accepted 11 May 2015
Available online 18 May 2015

Keywords:
Constructed wetland
Stormwater Management
Artificial neural network
Genetic programming
Velocity Flow Field
Acoustic Doppler Velocimeter

ABSTRACT

Monitoring the velocity field and stage variations in heterogeneous aquatic environments, such as constructed wetlands, is critical for understanding hydrodynamic patterns, nutrient removal capacity, and hydrographic impact on the wetland ecosystem. Obtaining low velocity measurements representative of the entire wetland system may be challenging, expensive, and even infeasible in some cases. Data-driven modeling techniques in the computational intelligence regime may provide fast predictions of the velocity field based on a handful of local measurements. They can be a convenient tool to visualize the general spatial and temporal distribution of flow magnitude and direction with reasonable accurancy in case regular hydraulic models suffer from insufficient baseline information and longer run time. In this paper, a comparison between two types of bio-inspired computational intelligence models including genetic programming (GP) and artificial neural network (ANN) models was implemented to estimate the velocity field within a constructed wetland (i.e., the Stormwater Treatment Area in South Florida) in the Everglades, Florida. Two different ANN-based models, including back propagation algorithm and extreme learning machine, were used. Model calibration and validation were driven by data collected from a local sensor network of Acoustic Doppler Velocimeters (ADVs) and weather stations. In general, the two ANN-based models outperformed the GP model in terms of several indices. Findings may improve the design and operation strategies for similar wetland systems.

1. Introduction

The Everglades, located in South Florida, is one of the largest tropical wetlands in the world. Historically, water flowed south through the Kissimmee River Basin into Lake Okeechobee and then continued south entering the headwaters of the Everglades. The water continued a southerly flow through the Everglades ecosystem, eventually discharging into the Florida Bay. Due to increasing anthropogenic activities, around 2830 km^2 of land situated south of Lake Okeechobee was declared as the Everglades Agricultural Area (EAA). As a consequence, much of the runoff through this area has been rerouted through drainage canals, altering local hydrodynamic properties. Human interference including increased fertilizer application and livestock has created a new source of nutrient loading entering EAA runoff, eventually flowing into the Everglades headwaters. Excessive nutrients entering the Everglades deteriorate the overall water quality, causing eutrophication in the wetland area. As a measure to prevent continuous deterioration of water quality due to nutrient loading, the South Florida Water Management District (SFWMD) developed a buffer zone of constructed wetlands named the Stormwater Treatment Areas (STAs) at the intersection between the Everglades Agricultural Area (EAA) and the Everglades headwaters. Nearly 230 km^2 to the south of Lake Okeechobee has already been converted to such constructed wetlands to polish the stormwater runoff and remove nutrients, especially with respect to total phosphorus. Among all the constructed wetland areas, STA-3/4 with an area of more than 64 km^2 is currently known as the largest constructed treatment wetland in the world dedicated to intercepting runoff entering the Everglades to reduce phosphorus loading (SFWMD, 2013).

Vegetation links wetland sediment to the water column, serving a unique role in wetland ecosystems and creating complex loops between their biotic and abiotic components. Macrophyte communities and associated periphyton act as main vegetation species in wetlands. Emergent and submerged macrophytes along with macrophytic algae play a prevalent role in phosphorus cycling and deposition depending on their physical, chemical, and biological interaction (SFWMD, 2013). Major portions of the wetlands in STAs are covered with submerged aquatic vegetation (SAV) and emergent aquatic vegetation (EAV) such as chara, hydrilla, cattail, southern naiad and algae species. Presence of such aquatic plant mixtures imposes complex directional patterns on surrounding water pathways, which affects the overall flow regime and velocity profile within the cell.

* Corresponding author. Tel.: +1 407 8231375; fax: +1 407 8233315.
E-mail address: nchang@ucf.edu (N.-B. Chang).

http://dx.doi.org/10.1016/j.ecoinf.2015.05.001
1574-9541/

Each STA is assigned a water quality-based effluent limit (WQBEL) that represents the numeric discharge limit to be applied to all permitted discharges from those STAs in the Everglades according to the United States Environmental Protection Agency (USEPA) (SFWMD, 2013). In 2012, the Florida Department of Environmental Protection (FDEP) (FDEP, 2012a,b) issued Everglades Forever Act (EFA) Watershed Permit (0311207) and the National Pollutant Discharge Elimination System (NPDES) Watershed Permit (FL0778451) for operating the Everglades STAs to achieve the WQBEL limit. Such WQBELs pledge that the effluent discharge will not cause or exceed the 10 microgram per liter ($\mu g \cdot L^{-1}$) (0.01 $mg \cdot L^{-1}$) total phosphorus (TP) (long-term geometric mean) established under 62–302.540, Florida Administrative Code (SFWMD, 2012). TP concentrations are measured in the STA areas at EPA-designated locations to monitor TP concentrations, providing a better balance of aquatic flora and fauna in the Everglades. Discharge samples collected from each STA are required to meet two limits: 1) TP concentration will not exceed 13 $\mu g \cdot L^{-1}$ (0.013 $mg \cdot L^{-1}$) as an annual flow-weighted mean for more than three out of five water years, and 2) TP concentration will not exceed 19 $\mu g \cdot L^{-1}$ (0.019 $mg \cdot L^{-1}$) as an annual flow-weighted mean in any water year. Several chemical agents have been used satisfactorily to remove TP in different wetland pilot-scale studies for post-treatment wetlands; however, the capital and operational cost, residual management, and incompatibility with the receiving area of the Everglades present concerns in implementing the full-scale size of such chemical treatment technology (SFWMD, 2002). Efficiency of those pilot-scale plants or Managed Wetland Treatment System (MWTS) were evaluated using iron (III) chloride and polyaluminum chloride (PACL). Addition of small doses (less than 5 $mg \cdot L^{-1}$) of aluminum chloride ($AlCl_3$) in the EAA runoff prior it entering the STAs were found to not reduce TP concentrations relative to the establishment of passive cattail marsh (SFWMD, 2013).

A well-designed constructed wetland should be able to uphold two major properties, namely hydraulic loading rate (HLR) and hydraulic retention time (HRT) (Kadlec and Knight, 1996). The homogeneous distribution of HRT within the treatment wetland is key to successful TP removal in a wetland system. Whereas HRT is strongly dependent on the flow velocity inside the treatment wetland, few studies have examined the distribution of flow velocity and direction in wetlands using modeling techniques. When limited data are available with respect to interactions between hydrological, meteorological, and physical processes, bio-inspired computational alogrithms, including data-driven models such as artificial neural network (ANN) and genetic programming (GP) models, may be efficient systems analysis tools in capturing local correlations and global trends of velocity distribution. ANN, a data-driven model inspired by the connections with brain science, has been successfully applied to help solve several water and environmental problems (Maier and Dandy, 2000), such as the effects of climate change on the discharge of dissolved organic carbon and nitrogen from river basins (Clair and Ehrman, 1996). For example, the effects of climate change on the discharge of dissolved organic carbon and nitrogen from a river basin were modeled and assessed using ANN (Clair and Ehrman, 1996). Of several ANN algorithms, the back propagation (BP) training algorithm was used as a learning algorithm managing more than 80% of previous relevant studies (Maier and Dandy, 2000). ANN models were also used to forecast water quality parameters such as salinity (Maier and Dandy, 1996), residual chlorine concentration in urban water systems (Rodriguez and Sérodes, 1998), and prediction of wind velocity at a particular location using a reference set of data (Bilgili et al., 2007); however, few studies were conducted for constructed wetlands to predict nutrient concentrations in effluent waters (Pastor et al., 2003; Tomenko et al., 2007; Akratos and Tsihrintzis, 2007).

Another type of relatively new neural network algorithm, extreme learning machinge (ELM), has received wide attention in relevant fields (Huang and Siew, 2005; Huang et al., 2004, 2006; Liang et al., 2006; Zhu et al., 2005). Compared to BP, ELM is a fast learning approach for training single-layer feed-forward neural networks (SLFN). In principle, ELM can analytically find suitable output weights by randomly assigning values to the input weights, without having to iteratively search for the optimized input and output weights as in traditional ANN models such as BP. As a result, the training process in the ELM is much faster than other ANN approaches. ELM is known to avoid over-fitting, a common issue embedded in many traditional ANN methods (Huang et al., 2006). Furthermore, an ELM with N hidden units (also called random units) can learn exactly N distinct observations with zero error (Huang et al., 2006).

In parallel with ANN models, genetic algorithm-based models are inspired by Darwinian process of natural selection and biological operation. A special kind of evolutionary computing algorithm, namely (GP), was first presented by Koza (Koza, 1992) as a development or extension of the genetic algorithm. GP models operate by preparing populations consisting of random numbers (also known as chromosomes). The fitness of each chromosome is then evaluated by comparing with a target value. Using genetic crossover, or sexual recombination number(s), computer programs are genetically reproduced with the help of Darwin's principle of survival of the fittest. Some examples of GP models previously used include hydraulic studies on velocity prediction in laboratory-scaled vegetated floodplains (Harris et al., 2003), the development of rainfall runoff relationship from synthetic data, the assessment of salt intrusion in estuarine environments, the analysis of flow over a flexible bed covered with vegetation (Babovic, 1996; Babovic and Abbott, 1997), the simulation of rainfall-runoff process (Whigham and Crapper, 2001), and the prediction of total organic carbon concentrations in a lake (Chang et al., 2014).

Correlations among hydrological features, meteorological factors, and flow behavior have not been well investigated for complex hydraulic systems such as constructed wetlands. Different hydrodynamic models using Manning's roughness coefficient or Chezy coefficient could capture the hydraulic behavior of these wetlands, but developing such hydrodynamic models requires numerous datasets for model calibration and validation, which are not always available for remote, biodiversified areas like constructed wetlands for model calibration and validation. Collecting such numerous number of datasets might not always be possible for many remote biodiversified areas like constructed wetlands. Data-driven models can be used as an alternative to hydrodynamic models to capture the nonlinear relationships between different hydraulic and hydrologic parameters associated with wetland ecosystems. Because no previous research has sufficiently established the application of data-driven models to predict hydraulic parameters of a wetland system, this study presents a unique attempt to compare the application of ANN and GP models to a wetland system. In addition to the artificial intelligence algorithms, the traditional stepwise regression model was also applied as a baseline to compare the hydraulic parameters for these wetland systems.

The main scientific contribution of this study is to confirm the application potential of different bio-inspired computational algorithms in predicting hydraulic parameters for a constructed wetland system. The objectives of this study were to: 1) test three preselected machine learning techniques to predict the hydraulic parameters, namely flow magnitude and flow direction, over different cells of the STA, 2) compare the performance between traditional statistical regression techniques and the preselected machine learning techniques, and 3) determine the operational issues in this ecological engineering system as a firm basis for future modification. With these study objectives, the following scientific questions were explored in this study: 1) Is it possible to predict the low flow velocity regime in a constructed wetland using computational intelligence models based on limited local measurements via a sensor network? 2) Among different computational intelligence models, which model provides the most accurate estimation of low flow velocity regime for a constructed wetland system?