



Estimation of real-time N load in surface water using dynamic data-driven application system

Y. Ouyang^{a,*}, S.M. Luo^b, L.H. Cui^{b,c}, Q. Wang^d, J.E. Zhang^b

^a USDA Forest Service, CBHR, 100 Stone Blvd., Thompson Hall, Room 309, Mississippi State, MS 39762, USA

^b Ecological Agriculture Key Laboratory of the Ministry of Agriculture, South China Agricultural University, Guangzhou 510642, China

^c College of Natural Resources and Environmental Science, South China Agricultural University, Guangzhou 510642, China

^d Tropical Research and Education Center, and Department of Soil and Water Science, University of Florida, 18905 SW, 280 St, Homestead, FL 33031, USA

ARTICLE INFO

Article history:

Received 19 July 2010

Received in revised form

18 December 2010

Accepted 29 December 2010

Available online 2 February 2011

Keywords:

DDDAS

Nutrients

Real-time

River

Water quality

ABSTRACT

Agricultural, industrial, and urban activities are the major sources for eutrophication of surface water ecosystems. Currently, determination of nutrients in surface water is primarily accomplished by manually collecting samples for laboratory analysis, which requires at least 24 h. In other words, little to no effort has been devoted to monitoring real-time variations of nutrients in surface water ecosystems due to the lack of suitable and/or cost-effective wireless sensors. However, when considering human health or instantaneous outbreaks such as algal blooms, timely water-quality information is very critical. In this study, we developed a new paradigm of a dynamic data-driven application system (DDDAS) for estimating the real-time loads of nitrogen (N) in a surface water ecosystem. This DDDAS consisted of the following components: (1) a Visual Basic (VB) program for downloading US Geological Survey real-time chlorophyll and discharge data from the internet; (2) a STELLA model for evaluating real-time N loads based on the relationship between chlorophyll and N as well as on river discharge; (3) a batch file for linking the VB program and STELLA model; and (4) a Microsoft Windows Scheduled Task wizard for executing the model and displaying outputs on a computer screen at selected schedules. The DDDAS was validated using field measurements with a very good agreement prior to its applications. Results show that the real-time loads of TN (total N) and NO_x (nitrate and nitrite) varied from positive to negative with the maximums of 1727 kg/h TN and 118 kg/h NO_x and the minimums of –2483 kg/h TN and –168 kg/h NO_x at the selected site. The negative loads occurred because of the back flow of the river in the estuarine environment. Our study suggests that the DDDAS developed in this study was feasible for estimating the real-time variations of TN and NO_x in the surface water ecosystem.

Published by Elsevier B.V.

1. Introduction

Clean water is of vital importance for human societies and natural ecosystems. Agricultural, industrial, and urban activities are the major sources for contamination and eutrophication of rivers and lakes (Carpenter et al., 1998; David and Gentry, 2000; Dodds and Welch, 2000). The concentrations of biologically available nutrients in excess in surface water can lead to diverse problems such as toxic algal blooms, loss of oxygen, fish kills, loss of biodiversity, and loss of aquatic plant beds and coastal reefs. Nutrient enrichment in surface waters can also seriously degrade aquatic ecosystems and impair the use of water for drinking, industry, agriculture, and recreation and for other purposes. With an increased understand-

ing of the importance of drinking water quality to public health and raw water quality to terrestrial life, there is a greater need to assess surface water quality.

In the past, to determine surface water quality in a stream, it is necessary to manually collect samples and send them to a laboratory for analysis. These analytical methods require at least 24 h or longer. However, when the human health or other instantaneous outbreaks such as algal blooms are concern, timely water-quality information is required. Timely water-quality information also is useful for other many reasons, including assessment of total maximum daily loads and the effects of urbanization and agriculture on a water supply. In response to the need for timely and continuous water-quality information, the US Geological Survey (USGS) has begun using an innovative, continuous, real-time monitoring approach for many nation's streams (<http://waterdata.usgs.gov/nwis/rt>). These real-time monitoring water quality data normally include discharge, flow velocity, dis-

* Corresponding author.

E-mail address: youyang@fs.fed.us (Y. Ouyang).

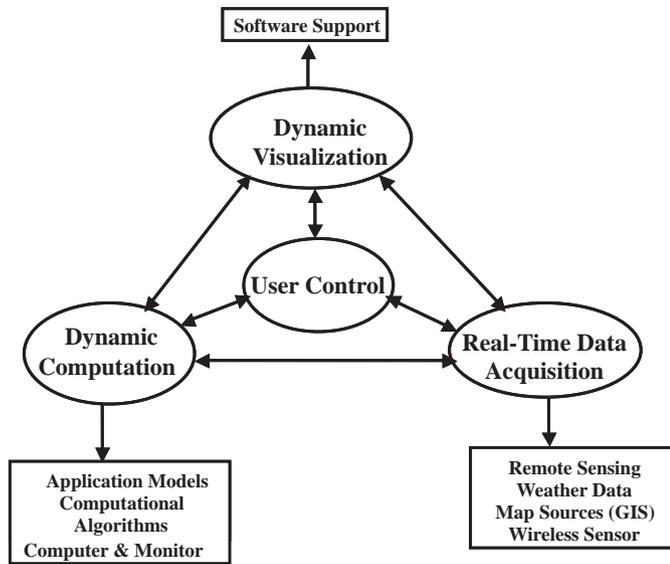


Fig. 1. A schematic diagram showing the basic concept of a dynamic data-driven application system redrawn after NSF (2000).

solved oxygen, pH, temperature, conductance, and chlorophyll. These data are valuable for monitoring surface water-quality indicators. However, there is currently very little activity to monitor the real-time variations of nutrients in surface waters due to the lack of suitable and/or cost-effective wireless sensors. Knowledge of real-time nutrient variations is critical to estimate surface water-quality status. Therefore, a need exists to develop a new paradigm for this purpose. To this end, a dynamic data-driven application system (DDDAS), which utilized the US Geological Survey (USGS) real-time chlorophyll *a* (Chl *a*) and river discharge data, a STELLA (Structural Thinking, Experiential Learning Laboratory with Animation) model for evaluation real-time variations of nitrogen (N), a VB.NET program, and the Windows interfaces, was developed in this study.

The DDDAS was probably first conceived by the US National Science Foundation around March 2000. Fig. 1 shows a basic concept of a DDDAS, which consists of the following four symbiotic components: user control, dynamic computation, real-time data acquisition, and dynamic visualization. A similar concept can also be found in NSF (2000), Douglas et al. (2004), Darema (2005) and Ouyang et al. (2007). Users control and interact with dynamic computation, real-time data acquisition, and dynamic visualization. Dynamic computation includes application models, computational algorithms, and all of the computing machines and their connections (e.g., computers and monitors). Real-time data acquisition involves the instantaneous data collections from remote sensing, climatic monitoring, GIS map sources, and wireless sensor measurements. Dynamic visualization includes supporting software and hardware for interactive visualization, which help users to control the system and make decisions.

When a DDDAS is launching, the dynamic computation infrastructure will start to run the application models and/or computational algorithms. Meanwhile, the real-time data acquisition infrastructure will start to collect the real-world data and inject them into the dynamic computation infrastructure for simulations. This DDDAS will have the ability to dynamically employ simulations to guide the real-time measurements, to determine when, where, and how it is best to gather additional data. In reverse, the DDDAS can also dynamically steer the simulations based on the real-time measurements. Such automatic steering of simulations and measurements with ability to switch between the two infrastructures can be envisioned through the dynamic visualiza-

tion infrastructure. The dynamic visualization infrastructure will be achieved through the software and hardware supports. Overall, all of the infrastructures are controlled and managed by the users. A specific example of a DDDAS applied to watershed contamination monitoring and predictions can be found in Ouyang et al. (2007).

Chl *a* is often used to estimate algal biomass, with blooms being predicted to occur when the Chl *a* concentration exceeds $40 \mu\text{g L}^{-1}$ (Stanley et al., 2003). During the last several decades, numerous studies have demonstrated a strong correlation among Chl *a*, total phosphorus (TP), and total nitrogen (TN) concentrations in north-temperate lake waters from around the world (Aizaki et al., 1981; Ahlgren, 1980; Sakamoto, 1966) and in Florida lakes (Huber et al., 1982; Canfield, 1983). Large- and small-scale experiments further suggested that P is a primary factor controlling algal growth, especially in northern lakes. Therefore, simple empirical TP–Chl *a* regression models (Dillion and Rigler, 1974; Jones and Bachmann, 1976) have been used to predict changes in Chl *a* concentrations in response to changes in TP concentrations. However, lakes surrounded by rich phosphate deposits and P-containing soils may be N limited. Existing equations using the P and Chl *a* correlation may inadequately estimate algal biomass under such circumstances. Canfield (1983) demonstrated that in Florida lakes, Chl *a* is significantly correlated with both TP and TN. The P is the limiting nutrient when the TP concentration is below $100 \mu\text{g L}^{-1}$, whereas the N is the limiting nutrient when the TP is above $100 \mu\text{g L}^{-1}$.

STELLA is a user-friendly and commercial software package for building a dynamic modeling system. It uses an iconographic interface to facilitate construction of dynamic system models. The key features of STELLA consist of the following four tools: (1) stocks, which are the state variables for accumulations. They collect whatever flows into and out of them; (2) flows, which are the exchange variables and control the arrival or the exchanges of information between the state variables; (3) converters, which are the auxiliary variables. These variables can be represented by constant values or by values depending on other variables, curves or functions of various categories; and (4) connectors, which are to connect among modeling features, variables, and elements. STELLA offers a practical way to dynamically visualize and communicate how complex systems and ideas really work (Isee Systems, 2006). STELLA has been widely used in biological, ecological, and environmental sciences (Hannon and Ruth, 1994; Costanza et al., 2002; Aassine and El Jai, 2002; Ouyang, 2008). An elaborate description of the STELLA package can be found in Isee Systems (2006).

The purpose of this study was to develop a DDDAS for indirectly estimating the real-time loads of N in a surface water ecosystem. Our specific objectives were to: (1) obtain the relationships between Chl *a* and total N (TN) as well as between Chl *a* and total Kjeldahl N (TKN) through linear regressions, using a long-term dataset from a regular (i.e., non real-time) surface water-quality monitoring station; (2) download the USGS real-time Chl *a* data from a monitoring station to a personal computer using a Windows-based VB.NET program; (3) develop a STELLA model for predicting the real-time loads of TN and NO_x (nitrate and nitrite) species in the surface water ecosystem based on the real-time Chl *a* data and the relationships obtained from Objective 1 as well as based on the river discharge data; (4) create a batch file for linking the VB.NET program and the STELLA model; (5) set up a Windows Scheduled Task wizard for implementing the DDDAS at given schedules; (6) validate the DDDAS for estimating real-time variations of N species using another independent dataset from the regular monitoring station; and (7) apply the DDDAS to forecast the real-time loads of N species in the surface water ecosystem.

It should be pointed out that the real-time monitoring station selected in this study was very close (<400 meters in distance) to the regular monitoring station in order to minimize the sample

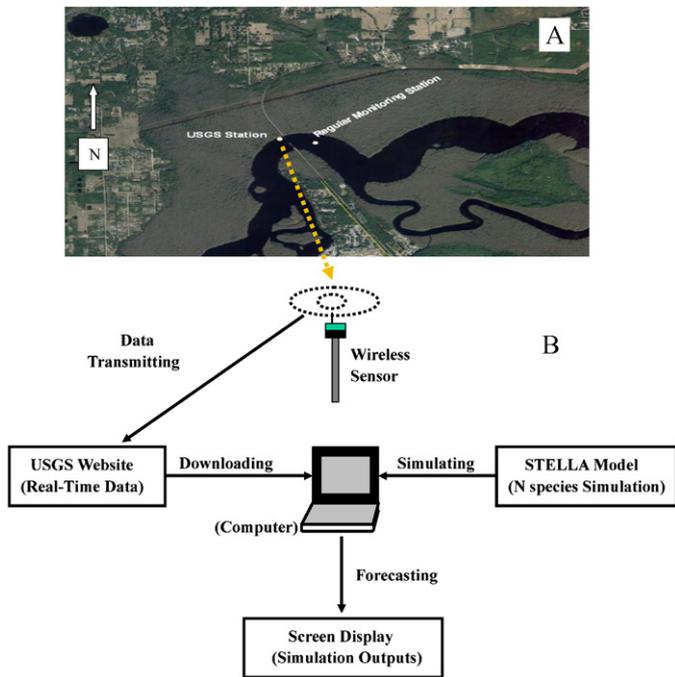


Fig. 2. (A) Location of the USGS real-time and regular monitoring stations near Satsuma, Putnam County, FL, USA. The distance between the two stations is <400 meters. (B) A schematic diagram showing a DDDAS framework for estimating real-time variations of N in a surface water ecosystem.

variations. As stated above, most of the USGS real-time monitoring stations do not measure nutrients in surface water ecosystems due to the lack of suitable and/or cost-effective wireless sensors. Therefore, it is impossible to directly estimate the real-time loads of N species based on the USGS real-time monitoring stations. The real-time loads of NO_x in the surface water ecosystem were calculated based on the real-time loads of TN and TKN that were obtained from Objective 3. The Windows Scheduled Task wizard employed in Objective 5 was used to guide the DDDAS on when to download the data, perform the STELLA simulation, display the simulations on computer screen, and end the real-time forecasting.

2. Materials and methods

A schematic diagram for a DDDAS in estimating the real-time loads of N species pertaining to this study is given in Fig. 2. This diagram shows the following five major components of the DDDAS: (1) a wireless sensor from a USGS real-time monitoring station; (2) a USGS real-time database website; (3) a STELLA model for simulating the real-time loads of N species in a surface water ecosystem; (4) a computer for downloading the real-time data and performing simulations; and (5) a screen monitor for displaying simulation outputs. The detailed descriptions of each component were presented below.

2.1. Data mining

The first step in developing the DDDAS is to select a study site (i.e., watershed) and a USGS monitoring station of interest from the USGS website within the watershed. This station should be very close to a regular (non real-time) monitoring station that has a long-term dataset for nutrients. In other words, these two monitoring stations should have the same representative for a watershed of interest. Once the real-time monitoring station is selected, a Windows-based computer

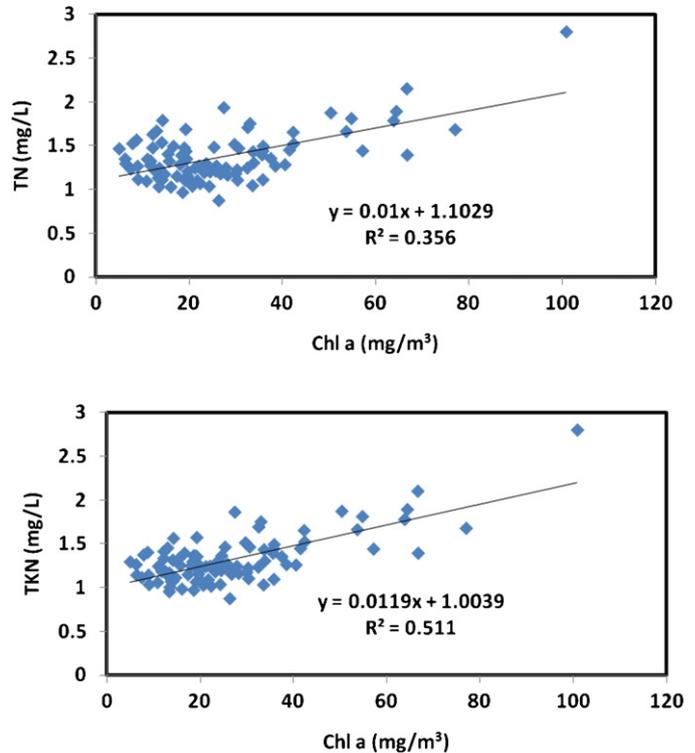


Fig. 3. Relationships of TN and TKN with Chl *a* obtained from field measured data.

program in Microsoft VB.NET needs to be constructed for simultaneously downloading the data to a personal computer. In this study, we selected a USGS real-time monitoring station #02244040 (Lat. $29^{\circ}35'46''$, long. $81^{\circ}41'00''$) located at the St. Johns River basin near Satsuma, Putnam County, FL, USA (http://waterdata.usgs.gov/fl/nwis/uv/?site_no=02244040&PARAMETER_cd=00400,00095,00010). In companion with this real-time station, there is a regular (non real-time) water quality monitoring station ($29^{\circ}35'43''$, $81^{\circ}40'45''$) located <400 meters east of the real-time monitoring station. This station is currently managed by the St. Johns River Water Management District (SJRWMD), FL. All sampling activities for this station were conducted in accordance with the SJRWMD and US Environmental Protection Agency's Standard Operating Procedures for the collection of water quality samples and field data. Both stations represent the same drainage area. However, the USGS station measured the real-time data on river flow characteristics and other water quality parameter such as Chl *a* but without nutrients, whereas the regular station collected most of the water quality parameters including nutrients but were not the real-time data and without river discharge. The nitrogen data collected during 1993–2003 from the regular monitoring station were used to obtain the relationships (Fig. 3) between Chl *a* (mg/m^3) and TN (mg/L) as well as between Chl *a* and TKN (mg/L) with the following linear regression equations:

$$\text{TN} = 0.01 \times \text{Chl } a + 1.01029$$

$$\times (R^2 = 0.356, p < 0.0019, \alpha = 0.05) \quad (1)$$

$$\text{TKN} = 0.0119 \times \text{Chl } a + 1.0039$$

$$\times (R^2 = 0.511, p < 0.00000184, \alpha = 0.05) \quad (2)$$

These two equations were used, respectively, to predict the real-time variations of TN and TKN concentrations based on the real-time variations of Chl *a* from the USGS real-time monitoring station.

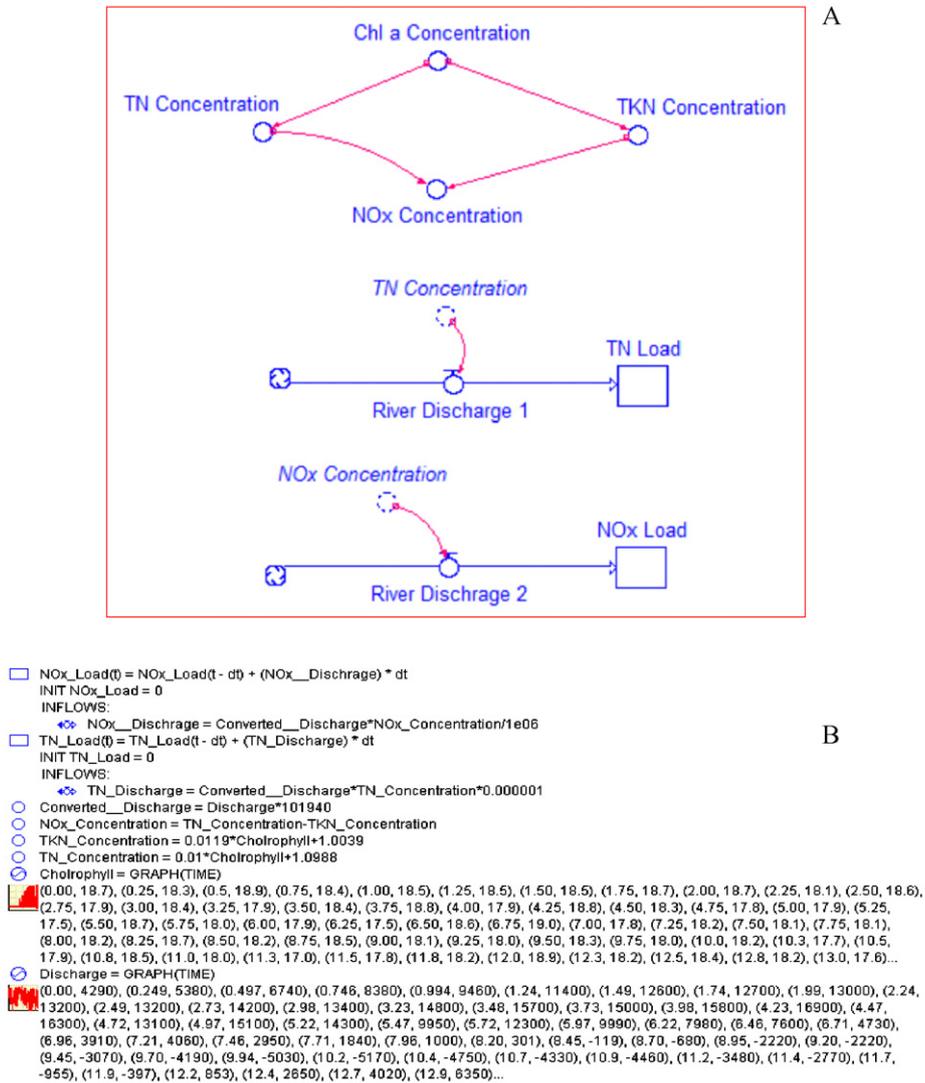


Fig. 4. STELLA modeling map (A) for evaluating N species loads and its code (B) showing the equations and graphic input data from the USGS station.

The real-time variations of NO_x (mg/L) were then calculated based on the following equation:

$$NO_x = TN - TKN \tag{3}$$

The real-time loads of TN and NO_x can be calculated using the following equations:

$$load_{TN} = 0.10194 \times discharge \times TN \tag{4}$$

$$load_{NO_x} = 0.10194 \times discharge \times NO_x \tag{5}$$

where *discharge* is the river discharge rate (ft³/s or 101,940 L/h) at the real-time monitoring station and the *load* denotes the masses of TN and NO_x loading from the station into the lower stream (kg/h).

2.2. STELLA model

The first step in the STELLA modeling processes was to develop a basic structure to capture the processes described above using STELLA. In Fig. 4, the rectangles are stocks that graphically represent the masses of nutrients. The flow symbols (represented by double lines with arrows and switches) represent the rates of nutrient discharges into or out of the stocks. The other variables are converters (represented by empty circles) that denote the rules or

conditions controlling the stocks and flows through the connectors (represented by single red lines with arrows).

As shown in Fig. 2, the model first received the real-time Chl *a* data from the USGS station; then calculated TN, TKN, and NO_x concentrations, respectively, using Eqs. (1)–(3); and finally estimated TN and NO_x loads, respectively, using Eqs. (4) and (5) in conjunction with the real-time river discharge data from the USGS station.

After the basic STELLA structure was developed, the second step was to assign the initial values for stocks, equations for flows, and input values for converters. The STELLA modeling code showing the equations and input parameter values are given in Fig. 4. This code was automatically generated with STELLA once its structure was established. It should also be noted that the STELLA software has an “Interface” module that can display simulation outputs instantaneously.

2.3. DDDAS framework

A batch file “RealTime.bat” was created by linking the following two executable files together: “usgs.exe” and “stella-N.exe”. The “usgs.exe” was written with Microsoft VB.NET for instantly downloading the real-time data every 30 min from the USGS website. This dataset was saved in a Microsoft Excel file. The “stella-N.exe”

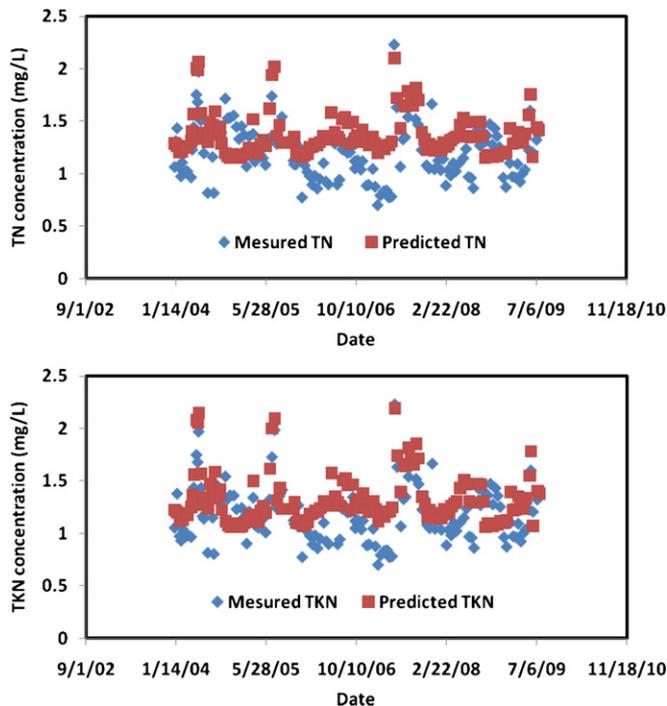


Fig. 5. Comparison of the DDDAS predictions with the field measurements for TN and TKN. The values of R^2 from the linear regression analysis were 0.692 and 0.748, respectively, for TN and TKN.

was composed with the STELLA package for modeling N loads and displaying the real-time predictions on a computer screen. The “stella-N.exe” read the Excel file for the real-time inputs of Chl *a* and river discharge when it was executed. A Microsoft Windows Scheduled Task wizard “RealTimeRun” in Windows XP was set up to include the “RealTime.bat” file and directed this bat file on when to begin and end running of the “usgs.exe” and “stella-N.exe” files as well as on the running time intervals.

In other words, the DDDAS developed in this study consisted of the following four files; (1) “usgs.exe”, (2) “stella-N.exe”, (3) “RealTime.bat”, and (4) “RealTimeRun”. To implement the DDDAS, users just need to click on the “RealTime.bat”.

3. Simulations

3.1. DDDAS validation

In order to apply the DDDAS for estimating real-time loads of N species in the surface water ecosystem, its applicability must be validated. The validation is a process of comparing the DDDAS predictions with the field observations within a given time period. In this study, an attempt was made to validate the DDDAS predictions using an independent set of the field observations collected from 2004 to 2009. Since no river discharge data were collected from the regular monitoring station, only the TN and TKN data from the regular monitoring station were used for validations.

Comparisons of the field measured and DDDAS predicted TN and TKN concentrations are shown in Fig. 5. The values of slope, intercept, R^2 , and p from the linear regression analysis were, respectively, 0.52, 0.75, 0.69, and <0.0000000006 for TN and were, respectively, 0.69, 0.52, 0.75, and <0.0000000005 for TKN. We, therefore, concluded that a fairly reasonable agreement between the field measurements and the DDDAS predictions was obtained.

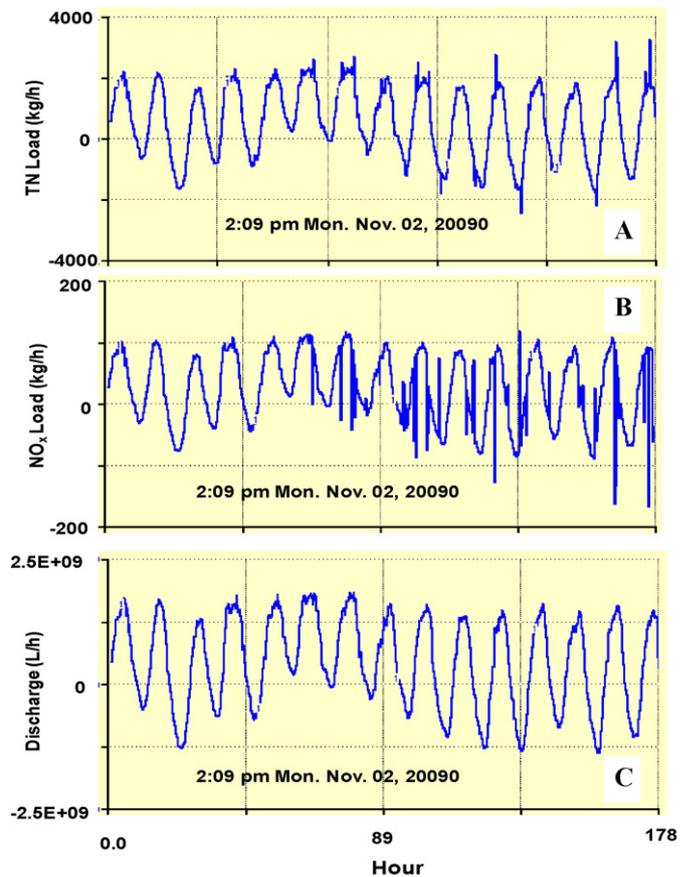


Fig. 6. Real time loads of TN and NO_x predicted from the DDDAS and real time variations of river discharge obtained from the USGS monitoring station from October 27 to 4 November 2, 2009.

3.2. DDDAS application

To obtain a better understanding of the real-time load of N in a surface water ecosystem, a forecasting (or simulation) scenario was performed in this study. In particular, this scenario investigated the real-time loads of TN and NO_x in responses to real-time variations of river discharge over a week period. Input values for the real-time river discharges and chl *a* contents at every 15 min were instantly downloaded from the USGS station (#02244040). The forecasting began on October 27, 2009 and ended on November 2, 2009. It should be pointed out that USGS only provides the most current 60 days’ real-time data for this station with an interval of 15 min. A week real-time data were selected in this scenario for the purpose of data storage efficiency and simplicity although it is very easy to modify the DDDAS for a 60-day period simulation.

Real-time variations of TN and NO_x loads in the surface water predicted from the DDDAS are shown in Fig. 6. It should be emphasized that although this figure demonstrated the variations of TN and NO_x loads for the entire simulation period (i.e., 7 days), in reality, the DDDAS was run every 15 min and the variations of TN and NO_x loads at that particular time were displayed immediately on the computer screen. The users can then estimate the surface water-quality status in a timely manner. The simulation ended at 2:09 pm on Monday, November 2, 2009.

Fig. 6 further reveals that loads of TN and NO_x varied from positive to negative with the maximums of 1727 kg/h TN and 118 kg/h NO_x and the minimums of -2483 kg/h TN and -168 kg/h NO_x. The negative loads implied that the TN and NO_x flowed back to the upstream, resulting from the negative (back) flow of river discharge

(Fig. 6C). The back flow of the river at the monitoring stations selected in this study was due to the tidal influence as the stations were located within the estuarine system. Comparison of Fig. 6A and B with Fig. 6C shows that effects of river discharge on TN and NO_x loads were profound, and the loads of TN and NO_x followed the same time-series pattern as that of the river discharge.

4. Summary

In this study, we had developed a DDDAS for forecasting real-time loads of TN and NO_x in a surface water ecosystem. Prior to its applications, the DDDAS was validated by field data with a reasonable agreement between the predictions and the measurements.

A forecasting scenario was chosen to demonstrate real-time loads of TN and NO_x in an estuarine surface water ecosystem. Results showed that river discharge had decisive effects on the real-time loads of TN and NO_x with the maximums of 1727 kg/h TN and 118 kg/h NO_x.

Our results further revealed that the DDDAS developed in this study was feasible for estimating the real-time variations of TN and NO_x in the surface water ecosystem. Further study is warranted to develop a DDDAS for scrutinizing the real-time loads of other water quality parameters such as phosphorus and organic carbon in surface water ecosystems.

It should be noted that the purpose for estimation of the real-time load of N in the river is not to monitor the algal bloom since the algal bloom can be better estimated from the USGS real-time Chl-*a* data. We estimated the real-time N load because high concentrations of N in a river could increase the biomass of aquatic plants, threaten the shallow groundwater quality, and affect terrestrial ecosystem. Additionally, for those areas around the world that use surface water as drinking water, the real-time monitoring of N load in surface water is very critical.

References

- Aassine, S., El Jai, M.C., 2002. Vegetation dynamics modelling: a method for coupling local and space dynamics. *Ecol. Model.* 154, 237–249.
- Ahlgren, I., 1980. A dilution model applied to a system of shallow eutrophic lakes after diversion of sewage effluents. *Arch. Hydrobiol.* 89, 17–32.
- Aizaki, M., Otsuki, A., Fukushima, T., Hosomi, M., Muraoka, K., 1981. Application of Carlson's trophic state index to Japanese lakes and relationships between the index and other parameters. *Verh. Int. Ver. Limnol.* 21, 675–681.
- Canfield, D.E., 1983. Prediction of chlorophyll a concentration in Florida lakes: the importance of P and N. *Water Resour. Bull.* 19, 255–262.
- Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N., Smith, V.H., 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecol. Appl.* 8, 559–568.
- Costanza, R., Voinov, A., Boumans, R., Maxwell, T., Villa, F., Voinov, H., Wainger, L., 2002. Integrated ecological economic modeling of the Patuxent river watershed. *Maryland Ecol. Monogr.* 72, 203–231.
- David, M.B., Gentry, L.E., 2000. Anthropogenic inputs of nitrogen and phosphorus and riverine export for Illinois. *USA J. Environ. Qual.* 29, 494–508.
- Darema, F., 2005. Dynamic data-driven applications systems: new capabilities for application Simulations and measurements. In: *Computational Science – ICCS 2005, Pt 2 Lecture Notes In Computer Science*, vol. 3515. Springer-Verlag, Berlin, pp. 610–615 (Heidelberger Platz 3, D-14197 Berlin, Germany).
- Dillion, P.J., Rigler, F.H., 1974. The phosphorus–chlorophyll relationship in lakes. *Limnol. Oceanogr.* 19, 767–773.
- Dodds, W.K., Welch, E., 2000. Establishing nutrient criteria in streams. *J. N. Am. Benthol. Soc.* 19, 186–196.
- Douglas, C.C., Shannon, C.E., Efiendiev, Y., Ewing, R., Ginting, V., Lazarov, R., Cole, M.J., Jones, G., Johnson, C.R., Simpson, J., 2004. A note on data-driven contaminant simulation. *Computational science – ICCS, Pt 3. In: Proceedings, Lecture Notes In Computer Science*, vol. 3038, pp. 701–708.
- Hannon, B., Ruth, M., 1994. *Dynamic Modeling*. Springer-Verlag, New York.
- Huber, W.C., Brezonik, P.L., Heaney, J.P., Dickinson, R.E., Preston, S.D., Dwornik, D.S., DeMaio, M.A., 1982. *A Classification of Florida lakes*. Science and Engineering Library, University of Florida, Gainesville, FL.
- Isee Systems, 2006. Technical Document for the iThink and STELLA Software., <http://www.iseesystems.com>.
- Jones, J.R., Bachmann, R.W., 1976. Prediction of phosphorus and chlorophyll levels in lakes. *J. Water Pollut. Control Fed.* 48, 2176–2182.
- NSF (National Science Foundation). 2000. Dynamic data driven application systems. 2000 NSF Workshop Report. http://www.nsf.gov/cise/cns/dddas/dd.das.work_shop.rprt.doc.
- Ouyang, Y., Zhang, J.E., Luo, S.M., 2007. Dynamic data driven application system: recent development and future perspective. *Ecol. Model.* 204, 1–8.
- Ouyang, Y., 2008. Modeling the mechanisms for uptake and translocation of dioxane in a soil-plant ecosystem with STELLA. *J. Contam. Hydrol.* 95, 17–29.
- Sakamoto, M., 1966. Primary production by the phytoplankton community in some Japanese lakes and its dependence on lake depth. *Arch. Hydrobiol.* 62, 1–28.
- Stanley, C.D., Clarke, R.A., McNeal, B.L., MacLeod, B.W., 2003. Relationship of Chlorophyll a Concentration to Seasonal Water Quality in Lake Manatee, Florida. Soil and Water Science Department, Florida Cooperative Extension Service, IFAS, University of Florida. SL-211, <http://edis.ifas.ufl.edu>.