Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 00 (2014) 000–000

Procedia Engineering

www.elsevier.com/locate/procedia

EUROSENSORS 2014, the XXVIII edition of the conference series

A Low Complexity Data Driven Model of Environmental Discharge Dynamics for Wireless Sensor Network Applications

Huma Zia, Nick Harris, Geoff Merrett*

Electronics and Computer Science, University of Southampton, Southampton, United Kingdom

Abstract

Poor water quality is a global concern, with agricultural practices the major contributor to reduced water quality with emissions of nutrient fluxes in to water systems. Using a collaborative framework to support catchment-scale water quality monitoring, control and management (WQMCM), individual sub-networks can learn and predict the impact of catchment events on their locality[1], allowing dynamic decision making for local irrigation strategies. Since resource constraints on network nodes (e.g. battery life, computing power etc) require a simplified predictive model, low-dimensional model parameters are derived from the existing National Resource Conservation Method (NRCS). An M5 decision tree algorithm is then used to develop predictive models for total discharge volume (Q), response start and duration (t_1 & t_d). Evaluation of these models demonstrates high accuracy (84-94%) even for a small training set of under 100 samples for Q and t_d. However, for t_1, 300 samples are required to give adequate performance.

Peer-review under responsibility of the scientific committee of Eurosensors 2014.

Keywords: Water Quality; Machine learning; M5 decision trees; NRCS Curve model

1. Introduction

Excessive or poorly timed application of irrigation water and fertilizers, coupled with inherent inefficiency of nutrient uptake by crops result in nutrient fluxes into the water system. However, it is challenging to make valid predictions about these outflows (what and when to expect). Due to the recent adoption of WSNs in precision

* Corresponding author. Tel.: +44(0) 23 8059 5000 ext. 23274.
E-mail address: nrh@ecs.soton.ac.uk

1877-7058
Peer-review under responsibility of the scientific committee of Eurosensors 2014.

agriculture, it is proposed that existing networked agricultural activities can be leveraged into an integrated mechanism by sharing information about discharges and predicting their impact [1, 2], allowing novel irrigation strategies to be implemented efficiently [3].

For discharge prediction, various hydrological models exist. One of the popular and simplified methods is the NRCS Curve number model [4]. However, at the time the NRCS method was developed, (pre-WSN), proxy parameters, average values or manual observations were used to represent land conditions. A WSN based system requires a simplified underlying physical model, based on fewer and, ideally, real-time field variables acquired autonomously. In the WQMCM framework [1], the output parameters of interest for discharge dynamics are Q, t_1 and t_d (See Figure 1).

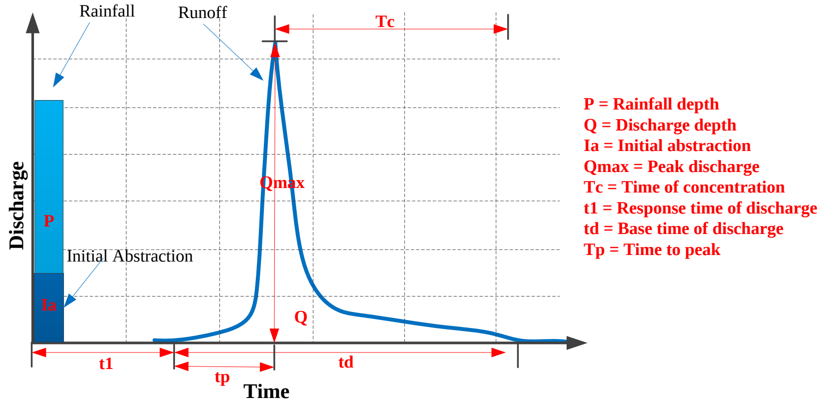

Figure 1: Plot of a hydrograph showing definitions of parameters

2. Model Simplification

Due to the availability of real field data from WSNs, empirical modelling using machine learning algorithms has become popular in hydrological forecasting [5]. In this paper, the NRCS model is used as a basis for deriving the simplified model parameters as illustrated in Figure 2 & Figure 3, resulting in a halving of the number of parameters required. This simplification is based on two steps; firstly the transient parameters from the NRCS model parameters are selected for each of the predictive models for Q, t_1 and t_d. This is because learning models are trained only on transient values. After this, the transient parameters are analyzed for likely improvements made possible by using available real field data from WSNs. For example, methods such as field imaging and signal attenuation methods have been used to determine the plant biomass autonomously [6]. This can give a measurement of crop stage. Similarly, various applications have used sensors to monitor soil moisture conditions of the field for precision irrigation [1]. Therefore, it is proposed to use actual soil moisture values instead of the 5-day rainfall index.

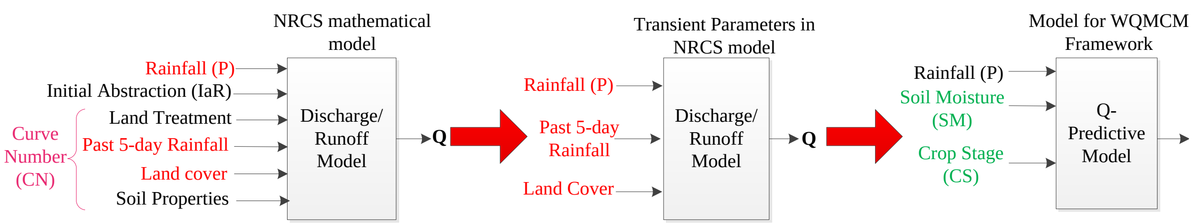

Figure 2: Model simplification for a Q-predictive model

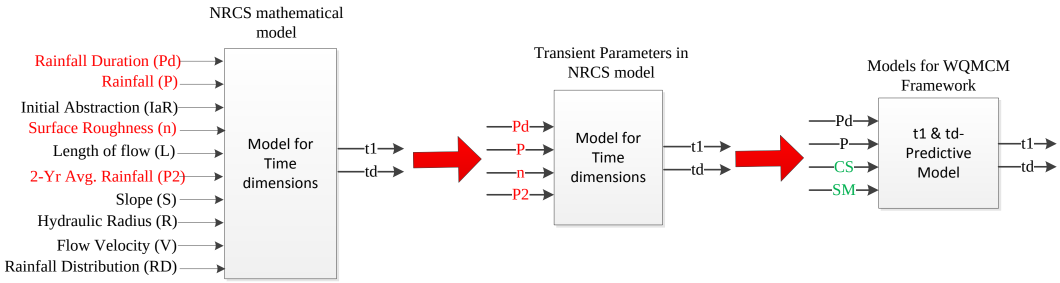

Figure 3: Model simplification for the t_1 and t_d-predictive models

3. Dataset

For training and testing the model, data is generated using a simulator based on the NRCS method [7], which was developed in Matlab. A combination of various event depths, field conditions and event duration is considered to generate two sets of data – one for the *Q* predictive model and the other for the t_1 and t_d predictive model. The obtained data set is then modified to substitute CN with the proposed simplified model parameters of CS and SM. To ensure robust evaluation of the model performance, the datasets are randomly sampled, in order to create training and testing subsets, respectively containing one-third and one-fourth of the available data.

4. Results and Discussion

Using these parameters, an M5-tree learning algorithm [8] generates the predictive models for *Q*, and t_1 and t_d. The prediction accuracy of the learned models is tested using RMSE (Root Mean Square Error), 10-fold cross validation (CVRMSE), Relative RMSE (RRSME) and R squared value (R2). A good value for RMSE and CVRMSE is stated as half of the standard deviation value for the output data. This comes out as 1.3 for *Q* and t_1, and 3.2 for t_d. Values of R2 and RRMSE can range between 0 and 1, where 1 means perfect forecasting. The value of RRMSE is represented as a percentage. The prediction results for these models show excellent match against the estimated output of the NRCS method (Figure 4). The *Q*-predictive and t_d-predictive model was tested to perform well even for a small training set of under 100 samples with 5.98% and 8% RRMSE respectively (Table 1). R2 for the two models is 0.984 and 0.99 respectively. However the t_1-predictive model required a minimum of 300 training samples to show reasonable performance (RRMSE=16.8%, R2=0.976).

Table 1: Performance of the predictive models based on various training sizes using M5 trees

	Q-Predictive Model (P, CS, SM)			**t1-Predictive Model (Pd, P, CS, SM)**			**td-Predictive Model (Pd, P, CS, SM)**		
Training set size	250	125	65	450	300	100	450	300	100
RMSE	0.159	0.234	0.317	0.239	0.318	0.825	0.2755	0.299	0.598
R2	0.998	0.997	0.984	0.985	0.976	0.835	0.997	0.977	0.991
CVRMSE	0.216	0.278	0.465	0.2935	0.381	1.042	0.3856	0.426	0.713
RRMSE	5.7%	7.5%	5.98%	16.1%	16.8%	27%	5%	6%	8.2%

Figure 4A also shows results from a model only using P, verifying that further simplification leads to poor results. RRMSE increases from 5.98% to 35%. Figures 4B and C also illustrate the performance of t_1 and t_d model when the model parameters of the *Q*-predictive model are used for training. In this case for t_1 and t_d, the plot shows very poor performance with RRMSE increasing from 16.8% to 65% and 8% to 98% respectively. This verifies the need for the inclusion of the P_d parameter for the prediction of t_1 and t_d.

In conclusion it can be seen that these results give confidence that the low complexity discharge prediction models described here can give excellent results when compared with standard methods, and further, the models are suitable for implementation on resource constrained wireless sensor networks.

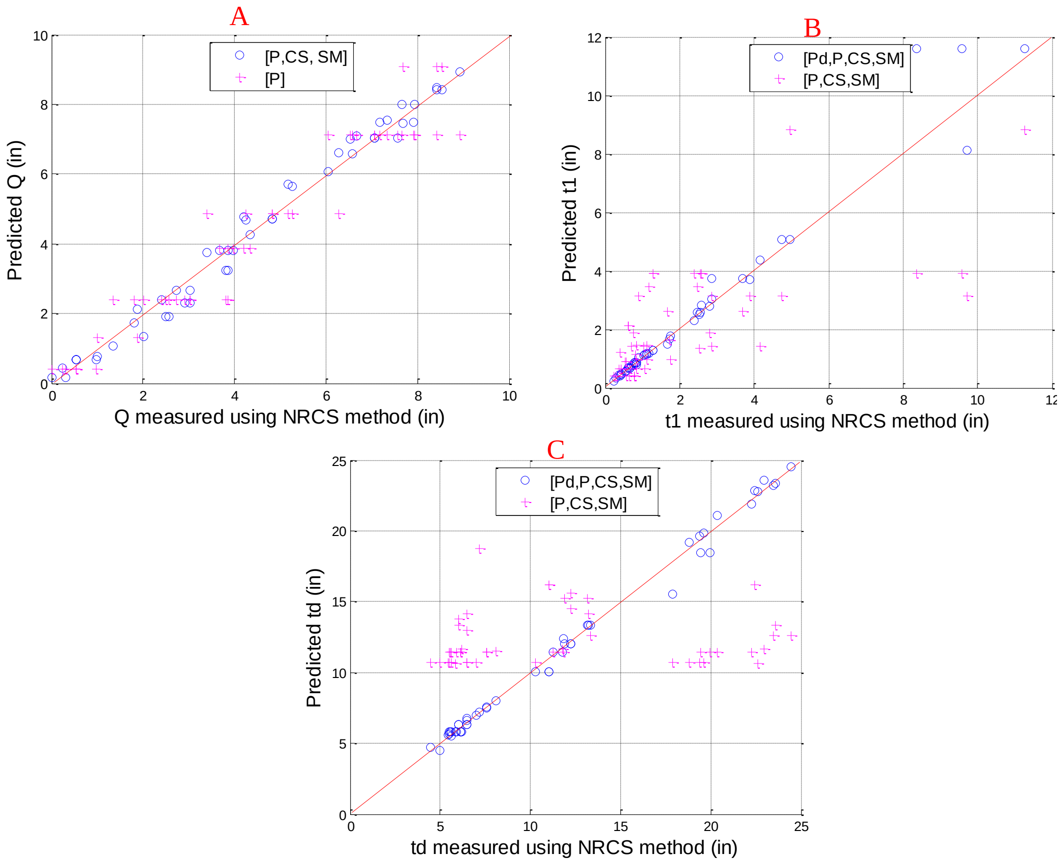

Figure 4: Plot of test data for, *A) Q*-predictive modes (100 samples), B) t_1-predictive model (300 samples), C) t_d-predictive model(100 samples)

References

[1] Zia, H., et al., *The impact of agricultural activities on water quality: A case for collaborative catchment-scale management using integrated wireless sensor networks.* Computers and Electronics in Agriculture, 2013. **96**(0): p. 126-138.

[2] Zia, H., N.R. Harris, and G.V. Merrett, *Collaborative Catchment-Scale Water Quality Management using Integrated Wireless Sensor Networks*, in *EGU General Assembly* 2013, Geophysical Research Abstracts: Vienna, Austria.

[3] Harper, H.H., *Impacts of Reuse Irrigation on Nutrient Loadings and Transport in Urbanized Drainage Basins*, 2012, Florida Stormwater Association, Environmental Research & Design, Inc.

[4] Hawkins, R.H., A.T. Hjelmfelt Jr, and A.W. Zevenbergen, *Runoff probability, storm depth, and curve numbers.* Jour. of Irrig. and Drainage Eng., 1985. **111**(4).

[5] Solomatine, D.P. and M.B. Siek, *Modular learning models in forecasting natural phenomena.* Neural Networks, 2006. **19**(2): p. 215-224.

[6] Vellidis, G., et al., *NDVI response of cotton to nitrogen application rates in Georgia.* Precision Agriculture, 2011: p. 359.

[7] Davis, T., *SCS Unit Hydrograph Convolution : Hydrograph Generation and Analysis Tool*, Matlab.

[8] Jekabsons, G. *M5PrimeLab: M5' regression tree and model tree toolbox for Matlab*. 2010; Available from: http://www.cs.rtu.lv/jekabsons/Files/M5PrimeLab.pdf.