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

UNIVERSITY OF Southampton

Huma Zia, Nick Harris and Geoff Merrett

School of Electronics and Computer Science, University of Southampton, Hants. SO17 1BJ, UK

Tel: +44 23-8059-3274 Fax: +44 23-8059-2931 Email: nrh@ecs.soton.ac.uk

Introduction

A potential application for wireless sensor networks is proactive precision agriculture. Such an application requires a simplified model of the environment, suitable for implementation on a resource constrained node to allow on-node prediction of events of interest. In this paper we describe a simplified dynamic model for predicting discharge events, driven by data generated within the network, which is suitable for implementation in a Wireless Sensor Network (WSN), allowing local decisions to be made.

Discharge Prediction

For discharge prediction, various hydrological models exist. One of the popular and simplified methods is the NRCS Curve number model [1]. 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. Fig. 1 shows a hydrograph, and defines the target output parameters Q, t_1 and t_d.

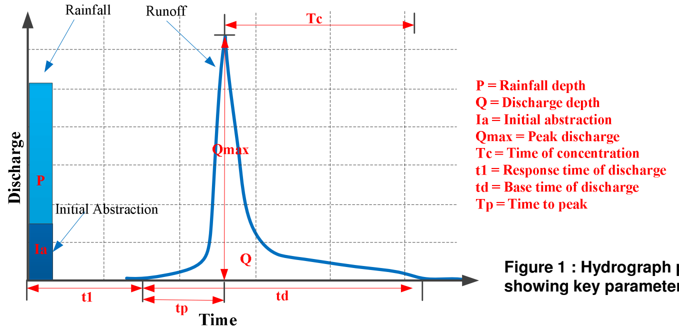

Figure 1 : Hydrograph plot, showing key parameters.

Model Simplification

Here the NRCS model is used as a basis, and simplified parameters are derived, halving the number required. There are two separate models – one predicts Q and the other predicts the time parameters.

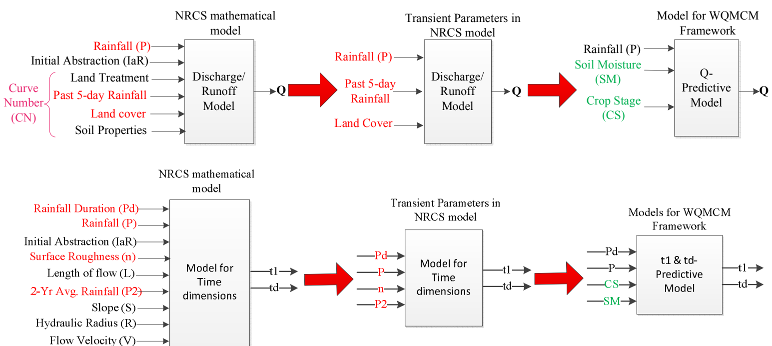

Figure 2 : Model simplification.

The simplification is based on two steps. The first is to identify the transient parameters in the NRCS model, as learning models are only trained on these. The second step then identifies how the real-time measuring capabilities of a WSN can best provide this data. For example, Past 5 day Rainfall can be replaced by Soil Moisture (SM), which can be dynamically measured.

Dataset

For training and testing the model, data is generated using a simulator based on the NRCS method, which was developed in Matlab [2]. 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 replace the resulting CN parameter with the proposed simplified CS and SM model parameters. 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.

[1] R.H. Hawkins, A.T. Hjelmfelt Jr, and A.W. Zevenbergen, *Runoff probability, storm depth, and curve numbers.* Journal of Irrigation and Drainage Engineering, 1985. **111**(4).
[2] T. Davis, *SCS Unit Hydrograph Convolution: Hydrograph Generation and Analysis Tool*, Matlab.

Results

Using these parameters, an M5-tree learning algorithm generates the predictive models for *Q*, t_1 and t_d. The prediction accuracy of the learnt models is validated 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. The prediction results for these models show excellent match against the estimated output of the NRCS method (Fig. 3). 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).

	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%

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

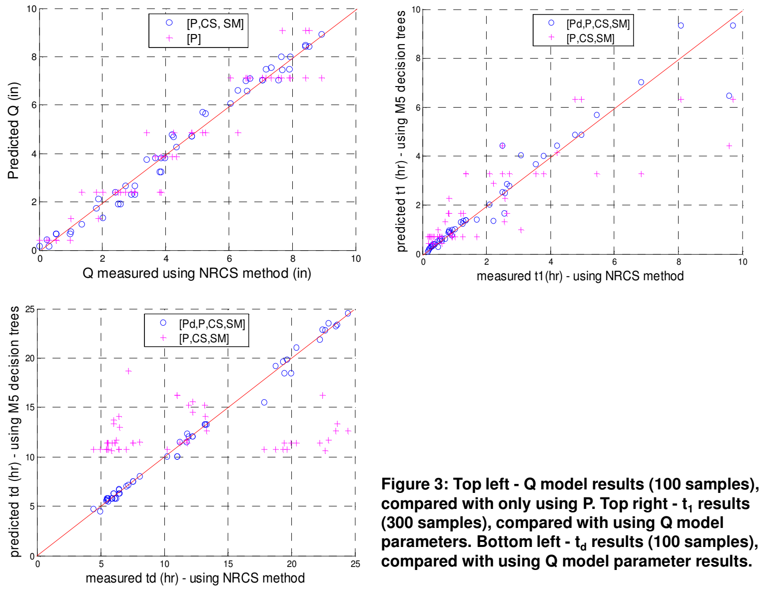

Figure 3: Top left - Q model results (100 samples), compared with only using P. Top right - t_1 results (300 samples), compared with using Q model parameters. Bottom left - t_d results (100 samples), compared with using Q model parameter results.

Figure 3a also shows results from a model using only P, verifying that further simplification leads to poor results. In this case, the RRMSE increases from 5.98% to 35%. Figures 3b and 3c illustrate the performance of t_1 and t_d model using the simplified parameters, and also when the P_d parameter is removed, i.e. the same as the Q model. 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.

Conclusions

In conclusion it can be seen that the results for these learning models give confidence that the low complexity discharge prediction models described here can give excellent results when compared with standard methods, and further, suitable parameters that play to the strengths of wireless sensor networks can be used, making the models advantageous for implementation on resource constrained wireless sensor networks.

Acknowledgements

The authors would like to thank Prof. Neil Coles and Prof. Mark Rivers of the University of Western Australia for invaluable discussions, and the World University Network for allowing such discussions to take place.

Eurosensors XXVIII, 7-10 September 2014, Brescia, Italy