Data-driven Low-Complexity Nitrate Loss Model utilizing Sensor Information – Towards Collaborative Farm Management with Wireless Sensor Networks

Huma Zia, Nick R. Harris, Geoff V. Merrett
Electronics and Computer Science
University of Southampton
Southampton, United Kingdom

Mark Rivers
Institute of Agriculture
University of Western Australia
Perth, Australia

***Abstract*—Excessive or poorly timed application of irrigation and fertilizers, coupled with the inherent inefficiency of nutrient uptake by crops result in nutrient fluxes into the water system. The ability to predict nutrient-rich discharges, in real time, can be very valuable to enable reuse mechanisms within farm systems. Wireless Sensor Networks offer an opportunity to monitor environmental systems with unprecedented temporal and spatial resolution, and we have previously proposed a novel framework to combine increasingly common local farm-scale sensor networks across a catchment to learn and predict (using predictive models) the impact of catchment events on their downstream environments, allowing dynamic decision making. Existing models require multiple parameters which are difficult to capture and this, coupled with constraints on network nodes (battery life, computing power, availability of sensors, etc.) makes it necessary to develop simplified models for deployment within the networks if they are to be utilized. The paper investigates a data-driven model for predicting daily total oxidized nitrate (TON) fluxes by assessing simplifications in model parameters and using only a year-long training data set. Data from a catchment in Ireland is used for training the model. Model simplification is investigated by abstracting details from an existing nitrate loss model. By using an M5 decision tree model on the training samples of the proposed parameters, results give R^2 as 0.92 and RRMSE as 0.26. The proposed novel model gives better results with fewer samples and simple parameters when compared to the traditional model. This shows promise for enabling real time nutrient control and management within the collaborative networked farm system.**

Keywords—nitrate losses, wireless sensor networks, agriculture, machine learning, M5 trees

I. Introduction

In order to increase crop yields, fertilizers rich in phosphorous, potassium and nitrogen are added to soils. However, agronomic nutrient recommendations are often far in excess of environmental levels [1]. Nitrogen (N) becomes a concern for water quality when N in the soil is converted into nitrate (NO_3^-). This form of N is not held tightly by soil particles, therefore, in a moist climate, nitrates remaining in the soil after the growing season can be lost to leaching and runoff into ground and surface water [2, 3]. In some cases, 30%-50% of the applied N is lost due to the combined impacts of over fertilization and irrigation runoff [4]. These substantial nutrient losses can have serious agronomic, economic and environmental implications [5]. Furthermore, the inherent inefficiency of nutrient uptake by crops (up to 70% for N) renders some nutrient losses inevitable [6]. This implies that adopting a mechanism which allows timely information (prediction) about nutrient outflows within the farm system can be a valuable strategy to manage these outflows before they reach regional waterways. In some intensive farming areas, farmers have begun to test their groundwater for nitrate concentrations and therefore change their nutrient budgets accordingly [7]. In another case, reuse of nitrate through the drainage water among golf courses was estimated to provide more than the annual nutrient requirements for the land [6]. Although this was based on manual sampling and unproven hypotheses it, nevertheless, illustrates the potential usefulness of improved in-farm water management.

With wireless sensor networks (WSNs) receiving considerable attention over the last decade, there now exists huge potential for leveraging small-scale networked agricultural activities among farms into an integrated mechanism by sharing information about discharges across networks and farms and, thus, enabling better impact prediction and pre-emptive management. However, there was no framework to investigate and implement such a mechanism until the authors proposed one, Water Quality Monitoring Control and Management (WQMCM) which utilizes collaboration among networks in a catchment to investigate and enable such a mechanism [8]. The basic model architecture comprises modules to enable individual networks to learn their environment by correlating neighbours' events with events within their own zone (neighbour-linking model), predict their impact in terms of discharges (Q-predictive model) and nutrient losses (N-predictive model), and then adapt the local monitoring and management strategy (Classification and Decision model). Work on the initial two models have been completed [9, 10]. This paper focuses on the development and evaluation of the nitrate (N) loss prediction model.

Two main approaches have previously been reported in the literature for nitrate loss modelling: process-based models, which use mathematical equations to conceptualize physical mechanism of nitrate loss; and data-driven approaches, in which the response variable is inferred from observational data [11, 12]. In process-based models, some models are more hydrology-oriented with less details about N-biogeochemical processes, such as MIKESHE and MODFLOW [13, 14], while others have focused more on the NO_3^- leaching processes, such as CENTURY and SOILN [15, 16]. In comparison to these, a modified De-nitrification-Decomposition (DNDC) model possesses an N=leaching module in addition to a complete set of N transformation processes [17-19]. However, the dependence of these models on acquiring numerous parameters (more than 20), the need for complicated calibration, and the tremendous computational burden involved in running the models makes wide-spread application complicated and difficult for sensor networks. Furthermore, the mis-calibration and over-parameterization results in a low predictive capability of the model [11]. In contrast, data-driven models have high prediction capability and require fewer parameters, which combine well with the computational burden of decision making [20]. Machine learning has most recently been adopted in the modelling of N losses. In this regard, a modelling framework was developed to calculate annual nitrous oxide flux and nitrate leaching by abstracting the complexity of the DNDC model [21]. The input parameters (11 variables) consisted of annual values related to N application, soil chemistry, and climatic conditions. Although this research effort reduced the number of parameters by half, it still required 8000 training samples based on annual values to get good results. Other works in this regard include [22, 23], which again reduced the number of model parameters but still relied on extensive chemistry data. Recently, regression methods were used in a 2 year study to simulate seasonal nitrate concentration dynamics in soil water extracted from 36 suction lysimeters in potato plots about seven to eight times each year . The model performed well, with R^2 of 0.95, however it used percentage of clay and soil depth (not easily measured) as well as other input parameters, and was based on sparse yearly samples [24].

From the relevant reviewed literature, it is apparent that existing modelling approaches have not been intended for predicting daily N losses within the farm system with the aim of enabling reutilization and alerts in real time by using WSNs. Furthermore, the reliance of existing models on acquiring chemical and geologic data, which either often requires grab sampling and laboratory analysis or very expensive equipment, limits wide-scale adoption of this technology for high resolution output. In addition, the strengths of a WSN deployment (fine spatial and temporal measurements of dynamic parameters) requires a simplified underlying physical model, and a simple machine learning model based on fewer and, ideally, real-time field parameters acquired autonomously and shareable across neighbour farms. Ideally, the model should be based on minimal training samples so that the model can be up and running soon after the deployment of the network.

In this paper we extend the concept of abstraction used in [21] to further simplify the model parameters with a view to eventual deployment within a WSN. This would enable wide-scale field management applications using WSNs without the need for complicated geo-chemical data. Furthermore, we explore the applicability of an M5 decision tree algorithm for nitrate loss prediction modelling based on the proposed simplified parameters. A year-long dataset consisting of daily values for N input, climatic conditions and N losses, obtained from a grassland catchment in Ireland are used for training and testing the model. Specifically, we adopt a three-step assessment procedure comprising: (i) abstraction of an optimized input parameter combination; (ii) random sampling of the observational dataset to ensure a robust evaluation of the model performance, and the use of 10-fold cross validation to avoid over fitting of the model, and; (iii) multi-criteria performance assessment of the proposed model.

II. METHOD

A. Data - Dripsey Catchment

A study was carried out by the University of Cork in the Dripsey catchment located in the south of Ireland. The one-year study (2002) was aimed at understanding the underlying processes of nutrient losses from soil to water bodies [25]. For the development of the TON–predictive model, precipitation (mm) and TON concentration (mg l^{-1}) for 2002 is used from a 17 ha sub-catchment of the Dripsey catchment. The data is available for research and education purposes at the Environmental Protection Agency (EPA) website [26]. The rest of the data regarding field conditions is extracted from catchment details available in the associated documentation [25]. For this sub-catchment, the cumulative rainfall for the year 2002 was 1812 mm, whereas the cumulative stream flow was measured as equivalent to 1206 mm at the outlet of the sub-catchment (as shown in *Fig. 1a*). The lower bound of the annual export of TON was estimated at 29 Kg N ha^{-1}, whereas the upper bound was 69 Kg N ha^{-1}. It can be clearly seen from

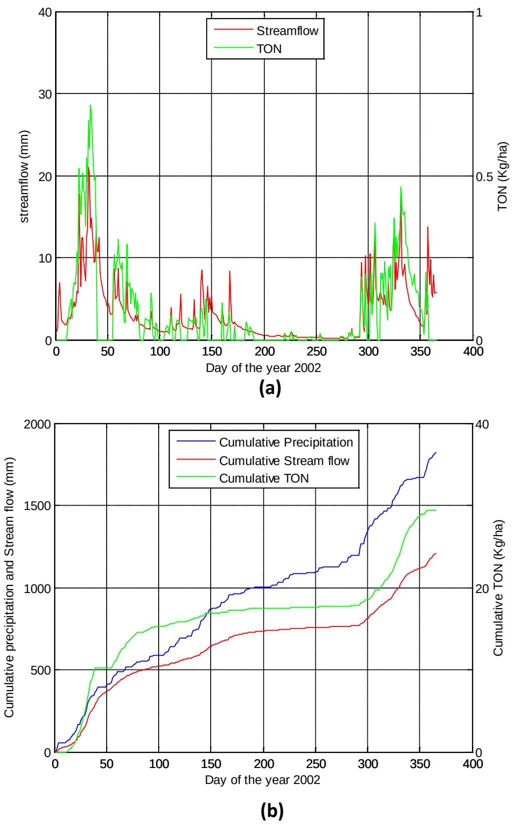

Fig. 1. : a) Plot of observed stream flow and TON and ; (b) cumulative precipitation, stream flow and TON losses, observed for the year 2002

Fig. 1b that the export of TON in the stream is strongly related to the sub-catchment stream flow.

B. *Learning Algorithm - Decision Tree Model*

After being widely used in hydrology modelling [27-29], data-driven modelling using machine learning is also being adopted to nutrient loss modelling [21]. Among the learning algorithms used in hydrology and nutrient modelling, artificial neural networks (ANN) and decision trees have been widely used [11, 21, 24]. One of the disadvantages of ANNs is that for a decision maker it is very difficult to analyses the structure of the resulting ANN and to relate it to the outputs. Furthermore, there are approaches to numerical prediction that are simpler and often reach accuracy comparable to that of ANNs. One example is M5 model trees [30], an extension of decision trees, which have been demonstrated as an alternative to ANNs [31].

Decision tree modelling is a method of approximating a target variable (output), with discrete values, from a given data set and represents the learned function in form of a decision tree [32], where each leaf contains the target values. To ensure good predicting accuracy, the number of tree leaves should be high; however, this increases the risk of over-fitting the observed data [33]. This can be resolved by replacing averaging in the tree leaves by fitting a linear regression function to the data and obtaining a continuous representation of the output [11]. This approach is known as M5 tree modelling, and was first introduced by Quinlan [30] and applied to hydrological modelling by Solomatine [31]. Model trees have higher predictive accuracy and are able to make predictions for values outside the training data range, which is not the case with regression trees [34]. However M5 decision trees have yet to be applied widely in nutrient modelling. To implement the M5 model trees, MatLab toolbox M5PrimeLab [35] is used.

C. *Model Evaluation Criteria*

To ensure a robust evaluation of the model performance, the data set was randomly partitioned (without replacement) into two groups: 75% of the observations were used for training the model while the remaining 25% are used to validate the model. When the available training data is small, in order to overcome the problem of over-fitting (meaning the model fits the training data but not unseen test data) and reduce the sensitivity of the model to the selected training set, a cross-validation technique allows reliable model validation [36]. 10-fold is the most commonly used cross validation, which is used in this paper, where data is partitioned into 10 subsets.

The prediction accuracy of the learned models is evaluated using multi-assessment criteria. The criteria considered are i) RMSE (Root Mean Square Error), which estimates the concentration of the data around the fitted equation, ii) Mean Absolute error (MAE), which measures the average magnitude of the errors in a set of forecasts, without considering their direction, iii) Relative RMSE (RRSME), which is the ratio of the variance of the residuals to the variance of the target values themselves and, iv) Coefficient of determination (R^2), which shows goodness of fit, i.e., correlation between actual and predicted values. RMSE and MAE are scale dependent measures and have the same unit as the data. A good value for RMSE is stated as half of the standard deviation value for the output data [37]. Values of R^2 and RRMSE can range between 0 and 1, where 1 means perfect forecasting. In this paper, the value of RRMSE is represented as a percentage.

III. Nitrate Predictive Model

As discussed in the Introduction, the adoption of WSNs for nutrient management in general and the implementation of WQMCM framework specifically on a farm, require simplified predictive models based on fewer, and ideally, real-time field parameters acquired autonomously and shared by the neighbour farms.

In order to simplify the model parameters, we extend the model abstraction done by Villa-Vialaneix *et al.* [21]. In that work, the input parameters consisted of 11 variables which were themselves drawn from the dataset of DNDC model based on a preliminary sensitivity analysis and expert evaluation. Table 1 lists input parameters for the two models (columns 2 and 3) under various input categories. Based on this abstracted list of parameters, which still contains soil chemistry and N-source data, we further abstract this to get the simplified parameters (3rd level abstraction, column 5). This 3rd level abstraction is explained below.

For the input category of climatic conditions, we select precipitation only. Temperature is not selected because temperature readings are taken into account in a model to imply the rate of evapotranspiration, which when combined with other soil properties such as field capacity and soil texture imply the soil moisture conditions and the eventual discharge flux from the soil [17]. Since, with WSNs, it is now possible to measure soil moisture directly with small and cheap sensors (see references within Zia *et al.* [8]), the dependence on proxy parameters can be minimized. Therefore, in the category of soil properties, we propose using only soil moisture.

In the category of N input sources, we select only “N in fertilizers” and “N in manure” from the parameters listed in the end level abstraction. This is because of the ease of availability of this information compared with other parameters listed in this category (such as N from precipitation, plant residues and atmospheric fixation) which require laboratory analysis of soil samples and mathematical modelling [17]. Besides those two parameters, we propose to use two additional parameters – “days since last N application” and “cumulative N applied so far that year”. The reasons for this are that it has been found that high monthly exports of nitrate do not always coincide with large monthly inputs of nitrogen fertilizer [24, 25] and these data are easily recorded and captured in real farming situations. Therefore, additional information related to N application is needed to develop a better relationship between N inputs and N fluxes. This will be corroborated in the later sections by the sensitivity analysis of the considered dataset and the model evaluation.

In the management information category, we propose using none of the parameters suggested in the 2nd level abstraction done by Villa-Vialaneix *et al.* [21]. This is because the dataset was comprised of annual values for the parameters; therefore, using annual averages for these variables would possibly not

TABLE I. ABSTRACTION OF INPUT PARAMETERS FOR THE TON-PREDICTIVE MODEL USING THE TRADITIONAL BIO-GEOCHEMICAL MODELS

Input variable category	**Parameters used in existing DNDC model Li *et al.* [17]**	**Parameters used by Villa-Vialaneix[21] *2nd level Abstraction***	**Proposed parameters for TON-Predictive model *3rd level Abstraction***
Climatic conditions	Precipitation	Precipitation	Precipitation
	Temperature	Temperature	
Soil Properties	Soil type		
	pH	pH	
	Redox		
	Carbon content	Carbon content	
	Bulk density	Bulk density	
	Clay content	Clay content	
	Temperature		
	denitrifying potential		
	Field capacity		Soil moisture
N input sources	Profile Mass		
	N in fertilizer	N in fertilizer	N in fertilizer
	N in manure	N in manure	N in manure
	N from precipitation	N from precipitation	
	N in plant residue	N in plant residue	
	N from fixation	N from fixation	
	N from mineralization		
			Total N applied
			Days since Last N application
Management Information	Crop cover		Crop cover
	Tillage		
	Crop rotation		
Additional Parameters			Day of the year

have contributed greatly to the model development. However, we have selected crop cover because of two reasons. Firstly, our model is assessing daily nitrate fluxes at the field scale in which vegetation cover can play an important role. Crop cover hinders outflows as well as impacting nutrient losses as nutrients are absorbed more in the initial stages of a crop [38]. Secondly, the availability of methods, using WSNs, enables autonomous monitoring of crop cover. For example, methods such as field imaging and signal attenuation methods have been used to determine the plant biomass autonomously [39]. An additional parameter, which was not used in either of the two previous models, is day of the year. Daily nitrate fluxes tend to have a clear (location-specific) trend over the year [24, 25] for small scale land areas irrespective of the timing of N application. Therefore, we propose using day of the year in the TON-predictive model to investigate its impact on prediction accuracy of the model.

Thus, the proposed parameters for the TON-predictive model (column 4 in Table I), abstracted from the two complex models, are precipitation, soil moisture, N in fertilizer, N in manure, total N applied, days since last N application, crop cover, and day of the year. All of these parameters are easily available. The three levels of abstraction from high to low complexity model parameters, along with model inputs and corresponding output parameters are shown in Figure 3.

IV. RESULTS AND DISCUSSION

The set of observations required for training the TON-predictive model was created after some pre-processing on the available dataset from the Dripsey catchment in Ireland. The data is available for research and educational purposes at the Environmental Protection Agency (EPA) website [26]. Despite many efforts, this is the only dataset that could be found freely with high temporal resolution data of TON losses for an entire year. From the available dataset we used data related to half hourly precipitation (mm) and TON losses (litres sec^{-1}) for the year 2002. The remaining parameters required for the TON predictive model were either obtained using a proxy value or were extracted from the information available in the documentation for this study [25].

Since the TON-predictive model is aimed at facilitating daily management decisions regarding nutrient loads (Kg N ha^{-1}), we convert the hourly values into daily loads. For soil moisture data, which is not available in the dataset (as sensor technology, still new at the time, was not adopted in this study), we use an alternative method. Instead of using the mathematical method suggested in the DNDC model [17] (because of the complexity of the required parameters) we use a proxy parameter - last-five-day-rainfall.

This proxy value has been widely used in hydrological models, such as in the NRCS curve number model, to represent soil moisture conditions , although there are questions about its accuracy and suitability [40]. Nevertheless, this is the best available proxy at present, and so offers a worst-case performance baseline. When real soil moisture readings become available to the model, performance should improve. Therefore, for each of the daily precipitation values, last-five-day-rainfall is computed. Using this value, moisture levels are determined according to the thresholds provided for growing and dormant seasons in the NRCS curve model. For example, in a dormant season, field conditions are considered dry, medium and wet respectively if rainfall depths are less than 13 mm, between 13 mm and 28 mm, and greater than 28 mm. Respective thresholds for rainfall depths are set for a growing season, which are higher than those for dormant season.

For obtaining crop cover data, information regarding growing stages of grass in catchment 1 was taken into account to obtain the estimates for crop coverage throughout the year. According to crop coverage values, crop levels are assigned such that fallow land is referred to as stage 1, coverage less than 20% is termed as stage 2, and coverage greater than 20% is assigned stage 3.

Similarly, information regarding N fertilizer and manure inputs to catchment 1 were extracted from a thesis based on the same project [25] and added to the dataset. Based on the N application rates and timings, cumulative-N application for each day and days-since-last-N-application are computed. The final dataset contains all the proposed attributes for the TON-predictive model. The daily mean TON for the dataset is 0.099 Kg ha^{-1}, 25th percentile is .040 Kg ha^{-1}, 75th percentile is 0.22 Kg ha^{-1}, and 90th percentile is 0.716 Kg ha^{-1}. The standard deviation is calculated as 0.156 Kg ha^{-1}. In the obtained dataset, we exclude instances with zero precipitation values which reduces the sample set to 200 event instances. A sensitivity analysis was done on the obtained dataset to analyse the correlation of various independent variables with TON losses. The values for Pearson coefficient of correlation are given in Table 2.

Using 75% of the pre-processed dataset as a training set, the M5 decision tree based TON-predictive model is generated by utilizing the M5 toolbox in MatLab. The generated model shows good performance with R^2 equivalent to 0.927, MAE as 0.024, RMSE as 0.040 and RRMSE as 26.4%. The 10 fold cross validated results indicate R^2 as 0.727, MAE as 0.043, RMSE as 0.065 and, RRMSE as 47.7%. As discussed earlier, RMSE values less than half of the standard deviation of the measured data may be considered appropriate for model evaluation. For the training dataset, this value is calculated as .078. Even for the 10-fold cross validated result for the model, the value for RRMSE falls well below this threshold. For testing the model, test samples are drawn from the remaining 25% of the dataset. Test results of the predicted TON values are plotted against measured TON as shown in Figure 5(a). The scatter plot shows a very good fit with R^2 equal to 0.91. To illustrate the difference between the predicted and measured TON curves, these values are plotted against day of the year. It is apparent that both curves overlap significantly, however the model seems to under predict in the last 50 days.

TABLE II. PEARSON CORRELATION COEFFICIENT FOR THE DATA FROM DRIPSEY CATCHMENT

Independent Parameters	**Acronym**	**Pearson (r) Correlation Coefficient with TON (Kg/ha)**
Day of the year	DY	-0.20
Precipitation (mm)	PPT	0.32
Soil Moisture	SM	0.71
Crop Stage	CS	-0.52
Last N Fertilizer Application (Kg/ha)	NF	0.12
Last N Slurry Application (Kg/ha)	NS	0.05
Days since last Fertilizer/Slurry Application	DNFS	0.31
Cumulative N applied so far this year (Kg/ha)	CumN	-0.27

In order to evaluate if the proposed model has acceptable (or comparable) performance, we compare its results with the meta-models developed by Villa-Vialaneix *et al.* [21]. For that research, various machine learning algorithms were used to develop the meta-models with different training set sizes. For performance measurement, only R^2 was evaluated, and cross validation was not done. The decision tree based meta-model resulted in an R^2 of 0.74 for 200 training samples. This indicates that the proposed TON model gives better performance for daily nitrate losses with R^2 equivalent to 0.92. The reason may possibly be attributed to the fact that the earlier model is for yearly estimates of nitrate losses which can overlook, by oversimplification, the complicated heterogeneous conditions through the year. In future a more rigorous prospective study with datasets obtained from different catchments is required to validate this further.

V. CONCLUSION

In this paper, we have successfully evaluated a nitrate (TON) predictive model for the proposed WQMCM framework by employing M5 decision tree learning approach. Simplified parameters related to climate, N input and field conditions were defined and used to avoid reliance on complex bio-geo-chemical parameters as used in the existing nitrate loss predictive models. The analysis was conducted on a nitrate loss dataset measured for a sub-catchment in Ireland. Results show that the proposed model provides better performance on the real dataset when compared to the existing nitrate loss models which required thousands of samples for training and complex parameters. For the generated model, R^2 is 0.92, RMSE is 0.04, and RRMSE is 26%. The results obtained for this simplified model reported in this paper show great promise for

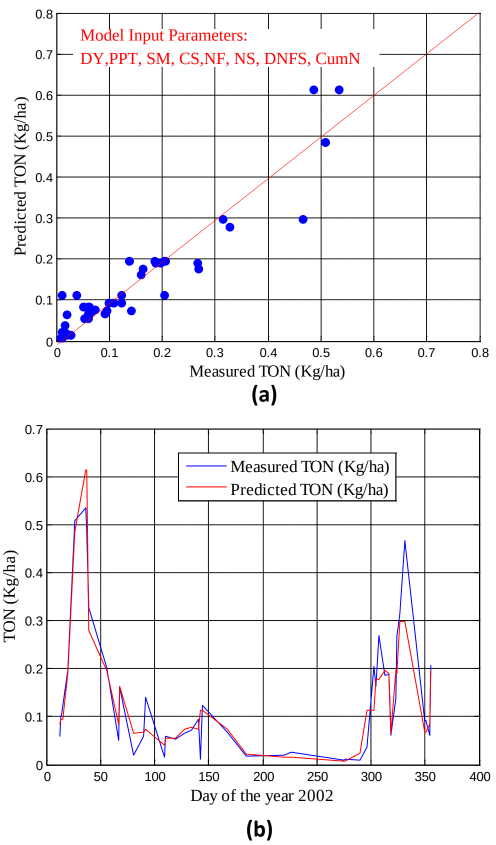

Fig. 2. (a) scatter plot of predicted TON against the measured TON; (b) plot of predicted and measured TON against day of the year

enabling real time nutrient control and management applications within the collaborative networked farm system.

REFERENCES

[1] D. K. Mueller, D. R. Helsel, and M. A. Kidd, *Nutrients in the nation's waters: too much of a good thing?*, Circular 1136 ed.: US Government Printing Office, Circular 1136, 1996.

[2] B. Vanlauwe, J. Wendt, J. Diels, G. Tian, F. Ishida, D. Keatinge, *et al.*, "Combined application of organic matter and fertilizer," 2001, pp. 247-279.

[3] L. A. McKergow, D. M. Weaver, I. P. Prosser, R. B. Grayson, and A. E. G. Reed, "Before and after riparian management: sediment and nutrient exports from a small agricultural catchment, Western Australia," *Journal of Hydrology,* vol. 270, pp. 253-272, 2003.

[4] X. Liu, X. Ju, F. Zhang, J. Pan, and P. Christie, "Nitrogen dynamics and budgets in a winter wheat-maize cropping system in the North China Plain," *Field Crops Research,* vol. 83, pp. 111-124, 2003.

[5] P. R. Ball and J. C. Ryden, "Nitrogen relationships in intensively managed temperate grasslands," in *Biological Processes and Soil Fertility*. vol. 11, J. Tinsley and J. F. Darbyshire, Eds., ed: Springer Netherlands, 1984, pp. 23-33.

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

[7] D. D. Adelman, "Simulation of irrigation reuse system nitrate losses and potential corn yield reductions," *Environmental Science & Policy,* vol. 3, pp. 213-217, 2000.

[8] H. Zia, N. R. Harris, G. V. Merrett, M. Rivers, and N. Coles, "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,* vol. 96, pp. 126-138, 2013.

[9] H. Zia, N. R. Harris, and G. V. Merrett, "Water Quality Monitoring, Control and Management (WQMCM) Framework using Collaborative Wireless Sensor Networks," presented at the 11th International Conference on Hydroinformatics HIC2014 New York City, USA, 2014.

[10] H. Zia, N. R. Harris, and G. V. Merrett, "Empirical Modelling and Simulation for Discharge Dynamics Enabling Catchment-Scale Water Quality Management," presented at the The 26th European Modeling & SImulation Symposium, Bordeaux, France, 2014.

[11] S. Galelli and A. Castelletti, "Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling," *Hydrology & Earth System Sciences Discussions,* vol. 10, 2013.

[12] D. Radcliffe and M. Cabrera, *Modeling Phosphorus in the Environment*: CRC Press, Boca Raton, USA, 2006.

[13] A. W. Harbaugh, E. R. Banta, M. C. Hill, and M. G. McDonald, *MODFLOW-2000, the US Geological Survey modular ground-water model: User guide to modularization concepts and the ground-water flow process*: US Geological Survey Reston, VA, USA, 2000.

[14] J. Refshaard, B. Storm, and V. Singh, "MIKE SHE," *Computer models of watershed hydrology.*, pp. 809-846, 1995.

[15] S. Liu, W. A. Reiners, M. Keller, and D. S. Schimel, "Simulation of nitrous oxide and nitric oxide emissions from tropical primary forests in the Costa Rican Atlantic Zone," *Environmental Modelling & Software,* vol. 15, pp. 727-743, 2000.

[16] H. Johnsson, L. Bergstrom, P.-E. Jansson, and K. Paustian, "Simulated nitrogen dynamics and losses in a layered agricultural soil," *Agriculture, Ecosystems & Environment,* vol. 18, pp. 333-356, 1987.

[17] C. Li, N. Farahbakhshazad, D. B. Jaynes, D. L. Dinnes, W. Salas, and D. McLaughlin, "Modeling nitrate leaching with a biogeochemical model modified based on observations in a row-crop field in Iowa," *Ecological modelling,* vol. 196, pp. 116-130, 2006.

[18] C. Li, S. Frolking, and T. A. Frolking, "A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity," *Journal of Geophysical Research: Atmospheres (1984–2012),* vol. 97, pp. 9759-9776, 1992.

[19] Y. Zhang, C. Li, C. C. Trettin, H. Li, and G. Sun, "An integrated model of soil, hydrology, and vegetation for carbon dynamics in wetland ecosystems," *Global Biogeochemical Cycles,* vol. 16, pp. 9-1-9-17, 2002.

[20] A. Castelletti, S. Galelli, M. Restelli, and R. Soncini-Sessa, "Tree-based reinforcement learning for optimal water reservoir operation," *Water Resources Research,* vol. 46, 2010.

[21] N. Villa-Vialaneix, M. Follador, M. Ratto, and A. Leip, "A comparison of eight metamodeling techniques for the simulation of N2O fluxes and N leaching from corn crops," *Environmental Modelling & Software,* vol. 34, pp. 51-66, 2012.

[22] F. Oehler, J. C. Rutherford, and G. Coco, "The use of machine learning algorithms to design a generalized simplified denitrification model," *Biogeosciences,* vol. 7, pp. 3311-3332, 2010.

[23] J. D. Piñeros Garcet, A. Ordoñez, J. Roosen, and M. Vanclooster, "Metamodelling: Theory, concepts and application to nitrate leaching modelling," *Ecological modelling,* vol. 193, pp. 629-644, 2006.

[24] J. Fortin, A. Morais, F. Anctil, and L. Parent, "Comparison of Machine Learning Regression Methods to Simulate NO 3 Flux in Soil Solution under Potato Crops," *Applied Mathematics,* vol. 2014, 2014.

[25] C. Lewis, "Phosphorus, nitrogen and suspended sediment loss from soil to water from agricultural grassland," NUI, 2003 at Department of Civil and Environmental Engineering, UCC., 2003.

[26] G. Keily. *Phosphorus, Nitrogen and Suspended Sediment loss from Soil to Water from Agricultural Grassland*. Available: http://erc.epa.ie/safer/resource?id=ad1f3acf-5035-102a-90c6-0593d266866d

[27] R. Wilby, R. Abrahart, and C. Dawson, "Detection of conceptual model rainfall—runoff processes inside an artificial neural network," *Hydrological Sciences Journal,* vol. 48, pp. 163-181, 2003.

[28] K. Rasouli, W. W. Hsieh, and A. J. Cannon, "Daily streamflow forecasting by machine learning methods with weather and climate inputs," *Journal of Hydrology,* vol. 414–415, pp. 284-293, 2012.

[29] D. Solomatine and A. Ostfeld, "Data-driven modelling: some past experiences and new approaches," *Journal of hydroinformatics,* vol. 10, pp. 3-22, 2008.

[30] J. R. Quinlan, "Learning with continuous classes," in *Proceedings of the 5th Australian joint Conference on Artificial Intelligence,* 1992, pp. 343-348.

[31] D. P. Solomatine and K. N. Dulal, "Model trees as an alternative to neural networks in rainfall—runoff modelling," Hydrological Sciences Journal, vol. 48, pp. 399-411, 2003.

[32] T. M. Mitchell, "Machine learning and data mining," *Communications of the ACM,* vol. 42, pp. 30-36, 1999.

[33] L. Breiman, "Bagging predictors," *Machine learning,* vol. 24, pp. 123-140, 1996.

[34] V. Kuzmanovski, "Integration of expert knowledge and predictive learning: Modelling water flows in agriculture " MS Thesis, Jožef Stefan International Postgraduate School, Ljubljana, Slovenia 2012.

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

[36] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI,* 1995, pp. 1137-1145.

[37] J. Singh, H. V. Knapp, J. Arnold, and M. Demissie, "Hydrological modeling of the iroquois river watershed using HSPF and SWAT1," *JAWRA Journal of the American Water Resources Association,* vol. 41, pp. 343-360, 2005.

[38] R. P. Udawatta, P. P. Motavalli, H. E. Garrett, and J. J. Krstansky, "Nitrogen losses in runoff from three adjacent agricultural watersheds with claypan soils," *Agriculture, Ecosystems & Environment,* vol. 117, pp. 39-48, 2006.

[39] G. Vellidis, H. Savelle, R. Ritchie, G. Harris, R. Hill, and H. Henry, "NDVI response of cotton to nitrogen application rates in Georgia," *Precision Agriculture,* p. 359, 2011.

[40] L. Fennessey and R. Hawkins, "The NRCS Curve Number, a New Look at an Old Tool," in *Proc. of Pennsylvania Stormwater Management Symp., Villanova Uni.*, 2001.