Science of the Total Environment 443 (2013) 539–548

Contents lists available at SciVerse ScienceDirect

Science of the Total Environment

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

Derivation of a fuzzy national phosphorus export model using 84 Irish catchments

Ahmed Nasr [a], Michael Bruen [b,*]

[a] School of Civil & Building Services Engineering, College Of Engineering & Built Environment, Dublin Institute of Technology, Bolton Street, Dublin 1, Ireland
[b] Centre for Water Resources Research, University College Dublin, Newstead Building, Richview, Belfield, Dublin 4, Ireland

HIGHLIGHTS

► Develops a new national phosphorus export model for agricultural catchments in Ireland
► Improves on earlier empirical phosphorus export models by using k-means clustering method for partitioning data
► Uses ANFIS model to predict annual average ortho-phosphorus concentrations using catchment characteristics
► Phosphorus desorption index (PDI) and runoff risk index (RRI) are essential predictors in the model.

ARTICLE INFO

Article history:
Received 29 August 2012
Received in revised form 15 October 2012
Accepted 15 October 2012
Available online 5 December 2012

Keywords:
Diffuse phosphorus pollution
Ortho-Phosphate
Empirical modelling approach
Fuzzy sub-sets
Neuro-fuzzy model
K-means fuzzy clustering

ABSTRACT

Implementation of appropriate management strategies to mitigate diffuse phosphorus (P) pollution at the catchment scale is vitally important for the sustainable development of water resources in Ireland. An important element in the process of implementing such strategies is the prediction of their impacts on P concentrations in a catchment using a reliable mathematical model. In this study, a state-of-the-art adaptive neuro-fuzzy inference system (ANFIS) has been used to develop a new national P model capable of estimating average annual ortho-P concentrations at un-gauged catchments. Data from 84 catchments dominated by diffuse P pollution were used in developing and testing the model. Six different split-sample scenarios were used to partition the total number of the catchments into two sets, one to calibrate and the other to validate the model. The k-means clustering algorithm was used to partition the sets into clusters of catchments with similar features. Then for each scenario and for each cluster case, 11 different models, each of which consists of a linear regression sub-model for each cluster, were formulated by using different input variables selected from among six spatially distributed variables including phosphorus desorption index (PDI), runoff risk index (RRI), geology (GEO), groundwater (GW), land use (LU), and soil (SO). The success of the new approach over the conventional lumped, empirical, modelling approach was evident from the improved results obtained for most of the cases. In addition the results highlighted the importance of using information on PDI and RRI as explanatory input variables to simulate the average annual ortho-P concentrations.

1. Introduction

Like many countries, Ireland has continually improved its agricultural output to meet local demand for food as well as for exports. Using the land in any intensive agricultural activity can cause adverse impacts on the environment unless appropriate measures to reduce these impacts are put in place such as water storage facilitates (e.g. De Martino et al., 2012; De Paola and Ranucci, 2012). Most of the reviews of water quality in Ireland revealed that diffuse transport of phosphorus (P) by surface and sub-surface flows from agriculture soil to the receiving waters is one of the major environmental problems (e.g. Lehane and O'Leary, 2012; McGarrigle et al., 2002; Toner et al., 2005). Soluble P in a form of ortho-P is readily available for plants and always leads to eutrophication in Irish Rivers and as a result there is a need for a catchment based management strategy that encapsulates all elements contributing to the loss of this form of P (Hutton et al., 2008). The EU Water Framework Directive (WFD) (EEC, 2000) provides the legal grounds required to develop and enforce such a management strategy. It mandates a thorough investigation to predict the impacts which will be produced by each possible management alternative.

The Three Rivers Project (MCOS, 2002) was one of the early and detailed studies conducted in Ireland with the aim of developing catchment based monitoring and management systems for the Boyne, Liffey and Suir catchments. A related project was the Lough Derg/Lough Ree Project (KMMP, 2001) which addressed the same objectives in the Three Rivers Project. In addition to the valuable management plans developed by these two projects, an important database required for modelling diffuse P loads was generated. Daly and Mills (2006) utilised some of this data to develop an empirical model described in Appendix I to estimate the annual average ortho-P concentrations from diffuse sources at the outlet of a catchment. Using a number of spatial variables derived from land use,

* Corresponding author. Tel.: +353 17163212; fax: +353 17163297.
E-mail address: michael.bruen@ucd.ie (M. Bruen).

0048-9697/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.scitotenv.2012.10.063

soil type, stocking densities, fertiliser P use and soil P levels Daly and Mills (2006) developed a series of linear regression models relating the different combinations of these variables with the average annual ortho-P concentration. From these models they chose the best empirical model as the one which provided the best calibration. In the analysis Daly and Mills (2006) used data from 84 different catchments in Ireland which have been selected on the basis of the following criteria: (i) nested catchments were avoided; and (ii) diffuse pollution represented the main contributor of phosphorus to the stream. Starting with all variables and using a backward-step regression procedure they eliminated the variables that had no significant effect on the linear regression model. In their final model only two variables were deemed significant and retained in the linear regression equation. These two variables were phosphorus desorption index (PDI) and runoff risk Index (RRI). Due to its simplicity and parsimony this type of empirical export coefficient modelling approach to predicting diffuse contaminant loads has been used widely (e.g. Daly and Coulter, 2000; Davies and Neal, 2007; Johnes, 1996; Lek et al., 1999; Meynendonckx et al., 2006; Su et al., 2006). Usually models of this type do not incorporate in their structure any representation of the actual physical processes involved in the mobilisation and transport of P but instead they seek to establish a numerical link between the contaminant load and the catchment characteristics which influence it.

Most of the simple empirical annual average P export coefficient models are multiple linear regression models that linearly relate the predictors, which are in most cases land use types, with the predictand representing the water quality parameter under consideration such as P soluble reactive concentration (e.g. McGuckin et al., 1999). However, the diffuse P transport process is in fact highly non-linear and the driver variables are in general not limited to land use types but they rather extend to include other predictors (e.g. soil P sorption capacity, residual soil P amounts, density of livestock, and climate). Therefore non-linear models are extremely needed in order to accurately express the non-linearity in the process. Furthermore the available observed P concentration data only covers a small number of years and this necessities the use of Monte Carlo simulation techniques to generate synthetic long time series that can be used in uncertainty analysis (e.g. McFarland and Hauck, 2001). Thus it will be useful if the candidate model can intrinsically account for the uncertainty in the data using the available actual data through an embedded modelling mechanism without the need for an external procedure to generate synthetic data used in uncertainty analysis. The non-linearity and the self-uncertainty modelling features can jointly be found in the fuzzy inference modelling systems (Jang, 1993). These models have been successfully implemented in a number of diffuse pollution modelling studies. For instance Schärer et al. (2006) used a fuzzy decision tree to estimate P export at a catchment scale and also Shrestha et al. (2007) modelled nitrate dynamics in a catchment using a hybrid deterministic–fuzzy rule based model.

The aim in this study is to broaden and strengthen the empirical modelling approach by employing an adaptive neuro-fuzzy inference system (Jang, 1993) to develop a new empirical P export model. In developing this model the available catchment data is first partitioned into a number of clusters based on similarities in their characteristics. A hypothesis is made here that there is a physical basis for the clustering and that if a separate P export model is calibrated for each cluster, a better prediction of diffuse P loads in a catchment would be obtained by combining the outputs of all the cluster models in proportion to the catchment's membership weighting for each cluster. This means that the models developed for each cluster contribute to the diffuse P loads prediction in a catchment depending on the degree by which this catchment belongs to the cluster. The newly developed model is intended to be used as a predictive tool at a catchment level across all the River Basin Districts in Ireland and also with the view that an analogous approach can be used in other countries. For direct comparison with previous models, the new model has been developed and tested with the same data used by Daly and Mills (2006) in their model.

2. Estimation of nutrients loads using catchment characteristics

The level of nutrients, including phosphorus (P) and nitrogen (N), in a stream is usually an indicator of the situation in its upland catchment. Therefore in situations where diffuse pollution is significant it is always possible to obtain some estimate of nutrient levels from empirical models conditioned on catchment characteristics. The particular catchment characteristics which result in a robust model may not be known in advance and hence a trial and error procedure is usually followed to determine the best catchment characteristics. The relationship between the nutrient loads and the catchment characteristics in the export coefficient models (e.g. McGuckin et al., 1999) is always described by a first order multiple linear regression model as follows:

$$L = b_0 + \sum_{k=1}^{nvar} b_k x_k \tag{1}$$

where

L	nutrient load;
x_k	the value of the k^{th} catchment characteristic;
$nvar$	total number of the catchment characteristics;
b_0	constant term of the linear regression model;
b_k	coefficient of the k^{th} catchment characteristic of the linear regression model.

The total number of terms in the linear regression model is equal to the total number of catchment characteristics which have been included in the model plus one. The constant term and the coefficients, (i.e. the model parameters) are estimated using the least squares parameter estimation method. To obtain reliable estimates for the parameters it is always recommended to use data from as many sites as possible. However it is also recommended to select sites from a homogenous region where similar catchment characteristics (e.g. phosphorus desorption index (PDI), runoff risk index (RRI) (see Appendix II for further explanation of PDI and RRI), soil types, land use types, geology, aquifer types) prevail so that the resulting model would be a better representation, but only of that region. Hence it is not advisable to use such a model in regions outside the one used in estimating its parameters.

Here a new approach has been developed to produce a class of model that can be more readily applied in heterogeneous regions. The approach is based on fuzzy inference systems already used extensively in hydrological and water quality modelling (e.g. Chen et al., 2006; Dixon, 2005; Haberlandt et al., 2002; Jacquin and Shamseldin, 2006; Marce et al., 2004; Nayak et al., 2004). These modelling systems integrate the outputs from a number of sub-models to estimate a single overall output. Each sub-model can be considered as representative of a specific region type where the catchment behaviour is assumed homogeneous. The data used in developing the model are for the same 84 catchments used by Daly and Mills (2006) to develop their national P model. Such a national model is a tool of extreme importance in managing diffuse P pollution at a catchment level in each River Basin District in Ireland. The newly developed model is aimed at providing an improved, albeit more complex, alternative national P model. The model is tested by using part of the data set to calibrate the model parameters and the remaining part to validate the performance of the resulting model.

3. New neuro-fuzzy national P export model

Using a single general equation to estimate diffuse P loads from catchment characteristics may work well for a single homogeneous region but may not give good predictions outside of this region. Thus its use for the whole of Ireland is questionable. The reason for this is the wide variability in the behaviour of the catchments used to derive the equation. If among those catchments there is a dominant cluster of