Ecological Engineering 94 (2016) 493–502

Contents lists available at ScienceDirect

Ecological Engineering

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

Artificial neural network ensembles and the design of performance-oriented riparian buffer strips for the filtering of nitrogen in agricultural catchments

F.M. Santin, R.V. da Silva, J.M.V. Grzybowski *

Federal University of Fronteira Sul, Rodovia RS 135 km 72, CEP 99.700-000, Erechim, RS, Brazil

ARTICLE INFO

Article history:
Received 11 February 2016
Received in revised form 25 May 2016
Accepted 11 June 2016
Available online 28 June 2016

Keywords:
Riparian buffer strip
Nitrogen removal
Artificial neural network ensembles

ABSTRACT

The design of riparian buffer strips (RBS) with adequate width to consistently serve conservation purposes is discussed. From the diverse ecological functions RBS, we concentrate on the filtering of agrochemicals, more especifically nitrogen loads due to uphill agriculture activity. In view of the numerous parameters that influence the RBS filtering properties, we propose and discuss a methodology to support the design of RBSs as a function of desired filtering properties. Towards this end, we use experimental data from previous studies on mean nitrogen influent, removal efficiency, soil type, vegetation density and mean RBS width to systematically train, validate and test about 6000 artificial neural networks (ANNs) of diverse architectures. The data from the original dataset are resampled using bootstrapping in order to provide a hundred training sets aimed at reducing the influence of stochastic variations in the training data upon the network outputs. We then compose an artificial neural network ensemble (ANNE) with the ANNs with best performances. The ensemble is aimed at delivering an estimate of RBS width when presented with unseen input data from the watershed. The relatively low values of test errors indicate that the neural networks retained relevant elements from the functional dependence between the input and output data. Sensitivity analysis is performed and indicates the vegetation cover type is the single most relevant input variable for the resulting model. We illustrate the application of the proposed methodology by presenting the study of the RBS from the Ligeiro River watershed, an important catchment in southern Brazil. The results from the ANNE are compared to those obtained by multiple linear regression, from which we conclude that the ANNE give consistently better results. At large, the results suggest that the existing buffer vegetation width is insufficient for filtering purposes and that the buffer width proposed in the Brazilian environmental law, thus suggesting that the quality of water supply for this catchment might be at risk.

1. Introduction

Scientific studies worldwide make it clear the importance of the RBS in the conservation of species and habitats, water courses, water quality and quantity (Phillips, 1989; Mander et al., 1997, 2005; EPA, 2005; Correll, 2005; Syversen, 2005; Sahu and Gu, 2009). It is well established that RBSs accomplish a number of essential conservation functions, such as filtering of surface and subsurface flow, mitigation of erosion in water bodies, improving microclimate, trapping suspending sediments, adhering and assimilating nutrients, binding dissolved pesticides, among others (Mander et al., 1997; Correll, 2005). The contamination of water by residues of nutrients from agriculture croplands is of special concern. This is due to the fact that nutrient inputs occur systematically along the years, thus having cumulative effects. From these nutrients, nitrates are a major concern due to their intensive application as fertilizers in agriculture croplands. Additionally, nitrate solubility makes it particularly prone to be transported by ground water flow. In this context, the importance of vegetation resides in the plant uptake and storage of nitrogen, which reduces nitrogen concentration in the discharge flow. As such, the RBF acts like a filter. Nevertheless, this requires that buffer strips have sufficient width and vegetation density, such that they can handle the nutrient load to which they are subjected.

The functions and importance of RBS are widely acknowledged. However, it is somehow cumbersome to systematically establish what the adequate RBS features would be for a given stretch of

* Corresponding author.
E-mail addresses: roberto.dasilva@uffs.edu.br (R.V. da Silva), jose.grzybowski@uffs.edu.br (J.M.V. Grzybowski).

http://dx.doi.org/10.1016/j.ecoleng.2016.06.008
0925-8574/

river. Several studies provided pieces of knowledge about this topic using different objectives and methodologies. In reference Sahu and Gu (2009), the authors applied the SWAT (Soil and Water Assessment Tool) to compare the effectiveness of buffer strips and contour strips. They concluded that the percentage of nitrate outflow decreases as the buffer strip increases and that the curve of filtering effectiveness vs. area gets flatter as area increases, thus revealing a nonlinear relation between buffer strip width and filtering effectiveness. In accordance with these results, the author in Syversen (2005) found by means of numerical simulation and experimentation that the specific retention per square meter is lower in buffer strips with 10 m as compared to 5 m. It was also found that a 10 meter buffer accomplishes significantly higher percent removal efficiency as compared to 5 m. This agrees with previous results that nitrate outflow decreases as the buffer strip increases (Syversen, 2005). Another important result regarding ground water nitrate removal that seems to broadly apply is that narrow but continuous buffer strips are more effective than wider but intermittent ones of comparable area (Correll, 2005). In Phillips (1989), the author applies the detention-time model and the RBDE (Riparian Buffer Delineation Equation) to evaluate the effectiveness of nitrate removal in runoff under the assumption that longer detention times contribute to reduce transport capacity. It was found that, although all buffer strips play a significant role in the conservation of water quality, their effectiveness is highly dependent on width. As the slope gradient and buffer width are the two most influential variables to buffer effectiveness in the filtration of runoff, the author concluded that buffer width is the single variable that can be more efficiently adjusted to improving buffer filtration effectiveness. In this context, the question that we pose is *how to objectively and straightforwardly design a buffer strip that can accomplish desired goals while considering particular characteristics of a stretch of river or sub-basin?*

During our investigation of an answer to this question, we evaluated the pros and cons of the application of empirical models, physical models, and artificial intelligence to similar problems. We found that methods based on simple regression, for example, might not adequately capture the multidimensional functional dependence between characteristics and function of RBSs, thus leading to models with low correlation coefficients (see reference (EPA, 2005) for results on simple regression). This problem can be partly solved using multiple regression, as more parameters are considered. In this case, however, the quality of the fitting relies too heavily upon the input variables at hand, as we show in the results section. Another class of models, the physical models based on balance equations, can be applied to describe nutrient removal. As a matter of fact, nutrient removal involves many coupled physical processes for which physical descriptions are available. Nevertheless, they sum up a large number of parameters (e.g., soil porosity, saturated hydraulic conductivity, water level, slope, plant nutrient intake, vegetation growth stage, meteorological data), which must be available or at least estimated. On the top of that, all these parameters have to be calibrated against observed data, which can be challenging or even not realizable due to the lack of complete and consistent datasets. In this context, neural networks provide a direct mapping from input to output data by learning the parameters of a mathematical description of the process for a given set of available data. This provides flexibility in the sense that the method itself does not require a fixed set of input data. Rather, it attempts to learn from the data that is available. A clear limitation of the approach is that the quality and consistency of the learning is tightly related to the quality of the data. Another limitation is that the resulting model is a black box, such that it is not in general expected to provide clear insights into the phenomenon under study. However, ANNs are long recognized to provide outstanding results in a vast number of fields, e.g., character recognition (Mao, 1998), web monitoring of medical home care devices (Gopinath and Reddy, 2000), breast cancer diagnosis (McLeod et al., 2014) and evolutionary strategy learning applied to game-playing (Chellapila and Fogel, 1999). Further, ANNs find applications in environmental problems associated to complex physical processes: groundwater level estimation and forecasting (Daliakopoulos, 2005; Lallahem et al., 2005), parameter estimation in ground water (Garcia and Shigidi, 2006), pesticide contamination in shallow groundwater (Sahoo et al., 2006) and nitrogen concentration (He et al., 2011). The advantages reported in the literature include satisfactory predictions, quick preliminary assessment and identification of predictive variables. The literature shows that ANNs can be applied as a support and assessment tools in the estimation of parameters and to the identification of optimal management options adapted to specific contexts or preferences. Further, Bayesian belief networks (BNNs) were recently proposed as an alternative methodology to assess the ecological services of multifunctional trees (McVittie et al., 2015; Barton et al., 2016), a somewhat similar problem.

While ANN were widely studied and demonstrated to give satisfactory response in diverse applications, artificial neural network ensembles (ANNEs) were shown to outperform individual neural networks in a number of situations (Namatame and Tsukamoto, 1993; Mao, 1998; Chellapila and Fogel, 1999; Gopinath and Reddy, 2000; Koch et al., 2013; Tosh and Ruxton, 2007). The main reason is that, in principle, the output errors in a committee of independently trained ANNs tend to cancel out, thus enhancing the identification of the true functional properties of the phenomenon under study. The use of ANNEs is aimed at improving the output accuracy by averaging the outputs of a set of individual neural networks that are individually accurate and feature highly uncorrelated errors (Namatame and Tsukamoto, 1993). The resamples in the training data are aimed at reducing the stochastic variation of the random sampling that would be likely to occur with a single training set (Tosh and Ruxton, 2007). On its turn, the different initial conditions for each network aims at leading to distinct minima over the parameter space. As such, the search for better approximations and improved generalization capabilities is more far-reaching.

In this paper, we propose a methodology based on ANNEs to support the design of RBS as a function of the desired nitrogen filtering effectiveness. Particularly, we seek an estimation of the riparian buffer width (BW) as a function of vegetation cover type (VCT), soil type (ST), mean nitrogen influent (MNI) and removal effectiveness (RE). Towards that end, we train, validate and test neural networks with six different architectures (numbers of input-hidden-output neurons): 4-2-1 to 4-7-1 for a hundred resamples of the original dataset and ten random sets of initial weight and bias conditions for each architecture. On this basis, we propose a highly reproducible methodology that provides a performance-oriented and systematic design support tool. It is aimed at enhancing and particularizing the process of establishing minimum configurations for RBS that will effectively accomplish their desired ecological conservation services. We show that, once composed, ANNEs could be readily applied in automated processes of estimating the width of RBS within entire catchment basins in a particularized manner for each stretch of river and on the basis of desired filtering properties.

2. Materials and methods

2.1. The Ligeiro River watershed

The Ligeiro River watershed is situated in the city of Erechim, Rio Grande do Sul state, Brazil, between 27° 39′ S and 27° 43′ S and 52° 14′ W and 52° 18′ W. It is within the Apuaê-Inhandava watershed, which in turn, is within the Uruguai River basin, an important river for hydroelectric power generation. The populations of Erechim city