

Artificial Neural Network Ensembles Applied to the Design of Performance-Oriented Riparian Buffer Zones: the Case of Nitrogen Filtering

Francele Michele Santin¹

PPGCTA, UFFS, Erechim, RS

Roberto Valmir da Silva²

PPGCTA, UFFS, Erechim, RS

José Mario Vicensi Grzybowski³

PPGCTA, UFFS, Erechim, RS

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 specifically nitrogen loads due to uphill agriculture activity. Towards this end, we use experimental data from previous studies on mean nitrogen influent, filtering efficiency, soil type, vegetation density and mean RBS width to systematically train, validate and test about 6,000 artificial neural networks (ANNs) of diverse architectures. The results show that the ensemble largely outperforms most individual networks. Further, the results indicate that the existing buffer vegetation width is insufficient for filtering purposes and that the buffer width proposed in the Brazilian environmental law is about 40% below the value advocated by the specialized ANNE for most sub-basins in the watershed.

1 Introduction

Scientific studies worldwide make it clear the importance of the riparian buffer strips (RBS) in the conservation of species, water courses, water quality and availability [1–7]. Although the functions and importance of RBS are widely acknowledged, it is somehow cumbersome to systematically establish what the adequate RBS features would be for a given stretch of river. Several studies provided pieces of knowledge about this topic using different objectives and methodologies. In reference [6], the authors used SWAT (Soil and Water Assessment Tool) to compare the effectiveness of buffer strips and contour strips, having concluded that the percentage of nitrate outflow decreases as the buffer strip increases. In accordance with these results, reference [7] found by means of numerical simulation and experimentation that the specific retention per square meter is lower in

¹francele.santin@gmail.com

²roberto.silva@uffs.edu.br

³jose.grzybowski@uffs.edu.br

buffer strips with 10 meters as compared to 5 meters, thus indicating that the initial area of the buffer strip does a more intensive filtering and that the curve of filtering effectiveness vs. buffer width flattens as the buffer width increases. It was also found that a 10 meter buffer accomplishes significantly higher percent removal efficiency, which confirms previous results that nitrate outflow decreases as the buffer strip increases [7]. 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 [1]. In [5], 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.

Towards a better understanding of the relations between characteristics and function of RBSs, the application of artificial intelligence techniques can be of value, since they are tailored to handle dense amount of data and learn from them. As a matter of fact, while ANN were widely studied and demonstrated to give satisfactory response in a wide range of situations, there are several applications in which Artificial Neural Network Ensembles (ANNEs) were shown to outperform individual neural networks [8–12]. The main reason is that, in principle, the output errors in an ensemble of independently trained ANNs tend to cancel out, thus enhancing error mean and variance.

In this paper, we investigate the application of ANNEs in the problem of estimating the required width of riparian buffer strips from its desired filtering properties. Particularly, we seek an estimation of the riparian buffer width as a function of vegetation type, soil type, nutrient concentration and filtering effectiveness. Towards that end, we train, validate and test neural networks with six different architectures 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 for the application of ANNE to handle the mapping of such relations and provide a systematic design-oriented account on the matter, aimed at enhancing and particularizing the process of establishing minimum configurations for RBS that will effectively accomplish their desired ecological conservation services. Once trained, validated and tested, ANNEs could be readily applied in automated processes of estimating riparian buffer widths within entire catchment basins, on the basis of desired filtering properties and accessible parameters.

2 Materials and methods

The region under study is the Ligeiro River watershed, located in the city of Erechim, Rio Grande do Sul state, Brazil, between $-27^{\circ}39'$ and $-27^{\circ}43'$ Latitude, and $-52^{\circ}14'$ and $-52^{\circ}18'$ Longitude. Rainfall in the watershed is distributed throughout the year with rainfall events in every month, thereby presenting a cumulative annual rainfall of about 2,200 mm. The Ligeiro River watershed is 21.18 km^2 in area, from which 4.94 km^2 are composed of as forests and 16.04 km^2 as crop land. According to the Brazil's Forest Code, the watershed should have 2.10 km^2 in area of riparian vegetation, however, it has only

0.94 km².

Two datasets were used thorough the research: the first one is a complete dataset from the literature [2] with 39 records including mean nitrogen influent (MNI), removal effectiveness (RE), vegetation cover type (VCT), soil type (ST) and buffer width (BW). This dataset was used for training, validation and test of the ANNs. The second dataset has 165 records corresponding to each of the sub-basins of the watershed. It is composed of experimental data, satellite data and empirical equations (MNI, VCT, ST). In this case, RE is a design choice ($RE = 90\%$) and BW is defined as output. To estimate the width of the RBS as a function of the parameters MNI, RE, VCT and ST data, we use these parameters as inputs to the network. VCT was input as a density parameter belonging to the interval $[0, 1]$ according to groups of land use classification performed in GRASS GIS 7.0 using the Landsat image of the watershed [13] and *in loco* validation of a dozen key points. Due to the absence of experimental data for the watershed, the MNI for each sub-basin was estimated on the basis characteristic nitrogen load from soybean croplands [14]. Towards that end, it was assumed the contribution of each sub-basin scales with its cropland area and concentrates equally along the length of the stretch of river to which it contributes for a constant depth. Further, since we are interested in the sub-superficial transport, we regard the amount of precipitation and estimate the share of runoff and evapotranspiration for the watershed. As a result, the MNI for the j th sub-basin was estimated as

$$MNI_j = \frac{1}{\mu} C \cdot \frac{A_j^{cropland}}{L_j^{river}} \cdot \frac{1}{P_j^{runoff}} \quad (1)$$

where MNI_j (ppm) is the nitrogen load for the j th sub-basin, C (mg/m^2) is the typical nitrogen load for the respective cropland, μ is the soil porosity, L_j^{river} (m^2) is the area of the longitudinal section of the river that receives contribution of the j th sub-basin, $A_j^{cropland}$ (m^2) and P_j^{runoff} (l/m^2) is runoff. Due to the river's morphology in the watershed, we consider all longitudinal sections to have unitary depth. Each ANN was designed to have 4 inputs (MNI, RE, VCT, ST) and 1 output (BW). The training, validation and tests sets were obtained as follows. From the original dataset, 15% of the records were randomly selected to compose the test set to be used for all the ANNs. The remaining records were resampled with substitution 100 times, in order to generate 100 datasets. Each of such datasets was divided into training and validation subsets having 70% and 15% of the total number of records of the original dataset. The training was performed using Backpropagation Algorithm, learning rate 0.7 and a maximum of 500 epochs. The sigmoid function was used as neuron activation function. The input and output data were normalized to the interval $[0, 1]$ to fit the activation function. We considered all architectures from 2 to 7 neurons in the hidden layer to satisfy the upper bound proposed in reference [15]. The parameters for each ANN were chosen during the validation process by using the early stopping criterion. After the training and validation and test of all the ANNs the test error vector was calculated. The ANNs were ranked according to the norm of the error vector and the ones with error below 35% were selected. The Pearson correlation coefficient was calculated from the test error vector for each pair of ANNs to compose a correlation matrix. From this matrix, the ANNs with correlation coefficients

Classification	Test error norm	Number of neurons in the hidden layer
1	0.18	6
2	0.22	5
3	0.24	7
4	0.29	7
5	0.30	6
6	0.31	7
7	0.32	6
8	0.32	3
9	0.33	3
10	0.34	7

Table 1: Training log for the best performing ANNs: the results show that several distinct network architectures were selected for the ensemble, most of them with relatively large hidden-layers, thus indicating that the relationship was captured more effectively by ANNs with more neurons.

below 0.5 (mild or no correlation) were selected to compose the ensemble. Following these criteria, the number of ANNs in the ensemble resulted to be $n = 10$. The ANNE was then applied in the study of the performance-oriented design of riparian buffer strips in the Ligeiro River watershed, considering 90% of removal effectiveness as a design criterion.

3 Results and discussion

3.1 The neural network ensemble

The training, validation and test of the ANNs took about two days to complete in a 8-core Xenon workstation. The training errors were found to monotonically decrease with the increase of epochs. This is not always the case of validation errors, which initially decrease to reach a minimum and tend to grow again. The point marked with an star (*) indicates the point of early stopping, from which the network parameters are collected. The resulting architectures and test errors are presented in table 1.

Further evaluation was performed to compare the performance of the network ensemble against that of the individual networks by randomly picking records from the original database. Test results indicated that the ANNE outperformed over 96% of the individual neural networks regarding error vector norm (figure 1) and mean error. Further, they consistently reduced the variance of the output errors relatively to individual neural networks (figure 2).

3.2 Performance-oriented design of RBSs

Buffer widths presented values distinguished in classes, despite the variability in the mean nitrogen influent for each sub-basin. A somewhat similar behavior was observed in reference [3], as the output of nitrogen and phosphorus into streams was comparably low despite the different input load. This illustrates the fact that the uphill part of the buffer

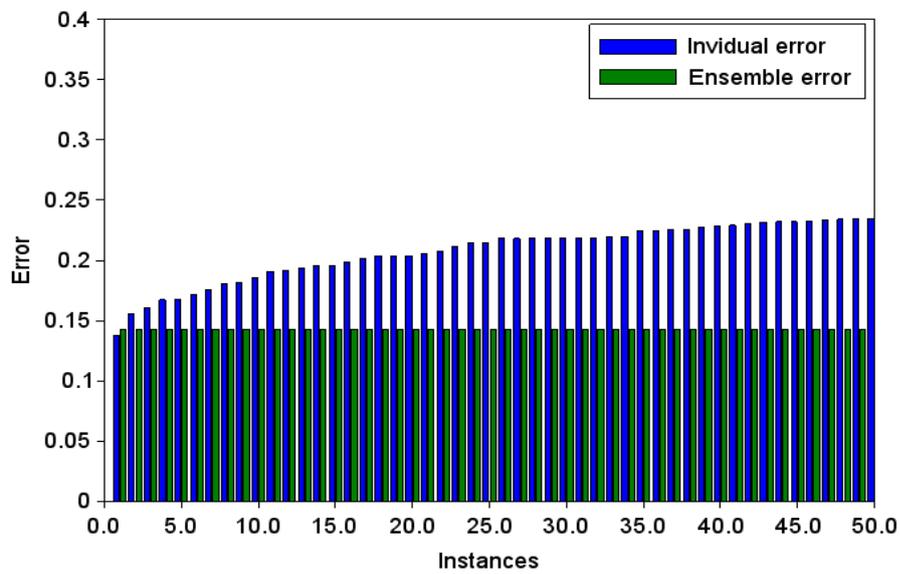


Figure 1: Individual error vs. ensemble error for the 50 networks with best error performance: the graph shows that the ensemble error magnitude is comparable to that of the best individual network.

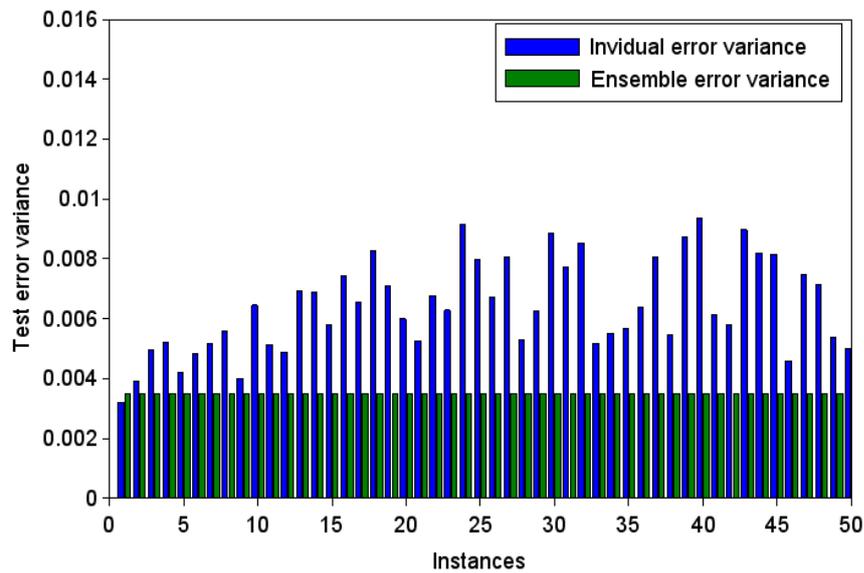


Figure 2: Individual variance error vs. ensemble variance error for the 50 networks with best error performance: the graph shows that the variance of the ensemble error is considerably smaller than that of the individual networks.

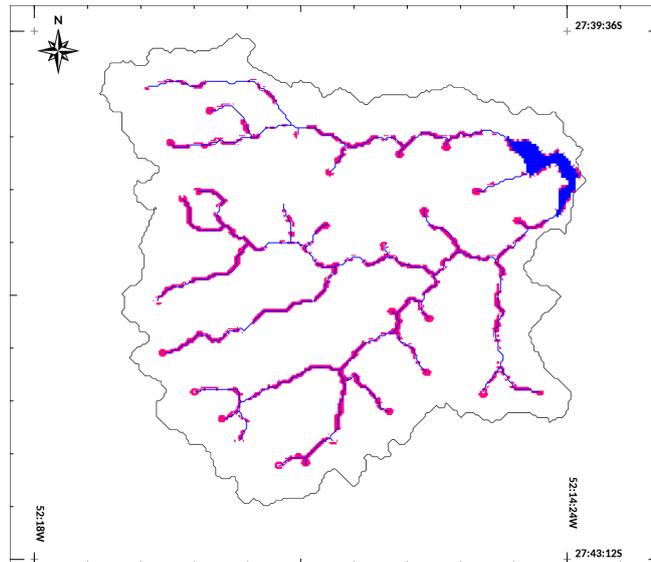


Figure 3: Missing RBS in the Ligeiro River watershed, obtained by means of comparing the ANNE output and the current situation obtained by land use classification in GRASS GIS 7.0: the estimated deficit amounts 1.97 km^2 of vegetated area according to the specialized ANNE and 1.16 km^2 according to the current environmental law.

vegetation retains a larger share of the nutrients, since the specific removal per meter is reported to decrease downhill. It is worth noting that the largest width value from the ANNE response, about 47 meters, is rather close to the 50 – 60 meter range indicated by the authors in [3]. In fact, the variability in ANNE buffer width was observed to depend more heavily upon the vegetation cover type (CT) and desired removal effectiveness (RE). Regarding CT, it was observed in [3] that higher nitrogen uptake observed in younger vegetation makes it more effective in nitrogen removal. This agrees with the results of the ANNE, since it was observed that grass-forest regions were identified as the areas less demanding of buffer width for the same filtering effectiveness. As one compares the results of the ensemble with the current buffer strips in the watershed it can be concluded that there is deficit of 1.97 km^2 of riparian vegetation (figure 3).

4 Final remarks

This study presented a primer on the application of neural network ensembles to the problem of estimating riparian buffer width on the basis of desired filtering properties. We turned our attention to nitrogen, whose presence nearby water bodies can severely affect water quality [2]. Based on the results, we can summarize that neural network ensembles can effectively capture the relations among the indicators related to the potential of mitigation of nutrient pollution.

References

- [1] D. L. Correll. Principles of planning and establishment of buffer zones. *Ecological Engineering*, 24:433–439, 2005.
- [2] EPA Environmental Protection Agency. *Riparian buffer width, vegetative cover and nitrogen removal effectiveness: a review of current science and regulations*. National Risk Management Research Laboratory, 2005.
- [3] U. Mander, V. Kuusemets, K. Lohmus, and T. Mairing. Efficiency and dimensioning of riparian buffer zones in agricultural catchments. *Ecological Engineering*, 8:299–324, 1997.
- [4] U. Mander, Y. Hayakawa, and V. Kuusemets. Purification processes, ecological functions, planning and design of riparian buffer zones in agricultural watersheds. *Ecological Engineering*, 24:421–432, 2005.
- [5] J. D. Phillips. Nonpoint source pollution control effectiveness of riparian forests along a coastal plain river. *Journal of Hydrology*, 110:221–237, 1989.
- [6] M. Sahu and R. R. Gu. Modeling the effects of riparian buffer zone and contour strips on stream water quality. *Ecological Engineering*, 35:1167–1177, 2009.
- [7] N. Syversen. Effect of buffer zones in the Nordic climate: the influence of width, amount of surface runoff, seasonal variation and vegetation type on retention efficiency for nutrient and particle runoff. *Ecological Engineering*, 24:483–490, 2005.
- [8] K. Chellapilla and D. B. Fogel. Evolving neural networks to play checkers without relying on expert knowledge. *IEEE Transactions on Neural Networks*, 10(6), 1999.
- [9] P. Gopinath and N. P. Reddy. Toward intelligent web monitoring: performance of committee neural network vs. single neural network. 2000.
- [10] J. Mao. A case study on bagging, boosting and basic ensembles of neural networks for ocr. volume 3, 1998.
- [11] A. Namatame and Y. Tsukamoto. Composite neural network models and their applications. volume 1, 1993.
- [12] C. R. Tosh and G. D. Ruxton. The need for stochastic replication of ecological neural networks. *Philosophical Transactions of the Royal Society B*, 362:455–460, 2007.
- [13] US Geological Survey. Landsat scene id: Lc82220792014348lgn00. 2014-12-14.
- [14] D. Harmel, S. Potter, P. Casebolt, K. Reckhow, C. Green, and R. Haney. Compilation of measured nutrient data for agricultural land uses in the united states. *Journal of the American Water Resources Association*, October:1163–1178, 2006.
- [15] R. Netch-Nielsen. Kolmogorov’s mapping neural network existence theorem. *First IEEE International Joint Conference on Neural Networks*, pages 11–14, 1987.