

A methodological approach for automatic weighting of variables for the planning of power transmission lines using artificial neural networks

ABSTRACT

Weighting variables for multicriteria analysis is crucial in power transmission line planning. This study proposes using an artificial neural network (ANN) called Neuralnet to automatically assign weights to variables. Inputs for training included social, environmental, cultural, and economic factors, such as land use, slope, and indigenous areas. The ANN was trained using 2000 and 200000 samples and used cross-entropy as the error metric. Results showed that 5 hidden neurons were sufficient to generate weights with a maximum success rate of 69% for the larger sample size. However, this success rate may be due to the low quality of samples used, as many layers were not considered in the construction of transmission lines used as positive samples. Overall, the study demonstrates the potential of using ANN to automate variable weighting in power transmission line planning, which can lead to more accurate decision-making.

KEYWORDS: Guideline Planning. Artificial Intelligence. Geoprocessing

Luiz Henrique Bossola

luiz.bossola@ufpr.br
orcid.org/0000-0001-6924-6576
Universidade Federal do Paraná
(UFPR), Curitiba, Paraná, Brasil.

Daniel Rodrigues dos Santos

daniel.rodrigues@ime.eb.br
orcid.org/0000-0001-7977-7426
Instituto Militar de Engenharia
(IME), Rio de Janeiro, Rio de
Janeiro, Brasil.

Kauê de Moraes Vestena

kauevestena@ufpr.br
orcid.org/0000-0003-1225-2371
Universidade Federal do Paraná
(UFPR), Curitiba, Paraná, Brasil.

Nathan Damas Antonio

nathandamas@ufpr.br
orcid.org/0000-0002-1469-2867
Universidade Federal do Paraná
(UFPR), Curitiba, Paraná, Brasil.

INTRODUCTION

As reported by the Brazilian national electric system operator (ANEEL, 2018), there are plans to build approximately 39.8 thousand kilometers of new power transmission lines (TLs) with a potential impact area of 199000 km² indirectly and 2000 km² directly. One of the most important factors in the construction of TLs is the planning stage (initial and experimental route definition), which involves prior knowledge of restrictive spatial and non-spatial factors concerning the project. Therefore, it can be considered as a multicriteria problem. Multicriteria analysis requires that an operator defines a set of weights for each layer involved in the decision process. Nevertheless, the criteria such as social, environmental, cultural, and economic impacts involved in the role process are subjectively judged by the operator (PERRAS, 2015; SPATH & SCOLOBIG, 2017; LAMADRID et. al, 2016; HEMMATI et. al, 2013). Thus, the initial and experimental route definitions are obviously much affected by the manual criteria used.

In the literature, many multicriteria analysis methods (MAMs) have been received highlighted such as the analytic hierarchy process - AHP (SAATY 1982), which structures the optimization with a hierarchy, analyzing the consistency of the weighting given by the analyst, the elimination et Choix Traudisant la Realité – ELECTRE (ROY, 1990), that relies on agreement and advantage measurement between variables, the preference ranking organization method enrichment evaluations - PROMETHEE (BRANS & VINCKE, 1985) that underlies on binary relationships, giving a proportional weight for according to its importance, the Electric Power Research Institute - EPRI (FRENCH ET, 2006) method, that works in a top-down scheme, with the planning starting from small scales, gradually advancing to small scales, and the environmental modeling methodology (CAMPOS, 2014) that uses a bayesian probabilistic approach, leveraging the “weights of evidence” automatically in a deterministic way.

Nowadays, with the surging of new generation technologies such as artificial intelligence (AI For example, the Artificial Neural Networks (ANN), they mimetic the Neural Networks (NN) from living beings, using simple processing units, like neurons, that can make connections, forming a network (KAMYAB, H. R.; ALIPOUR VARAKI, 2020; PECHANEC et. al., 2014). The main core of ANN is that it generally does not make any decisions based on explicit instructions or rules. In practice, it progressively learns how to recognize and generalize some patterns and relationships between inputs and resources (PECHANEC et. al., 2014).

The development of methods for the automatic planning of TLs, or TLGP - Transmission Lines Guideline Planning, has not been largely contemplated in the context of AI applications, the few found examples were YANG et. al (2022) which used an ant colony to optimize the process in the context of smart cities, the main finding was that the algorithm needs to be improved to understand the different altitudes of the crosslines; and GEORGESCU et. al. (1997) focused on using ANN to identify patterns of load infrastructure, with large performance upgrades when compared to traditional methods.

The present paper introduces a new methodology for planning TL, using ANN to establish the weights for data layers from public data.

The remainder of this paper is organized as follows: in Section 2 we will present an overview of the proposed method. In Section 3 we discuss the results

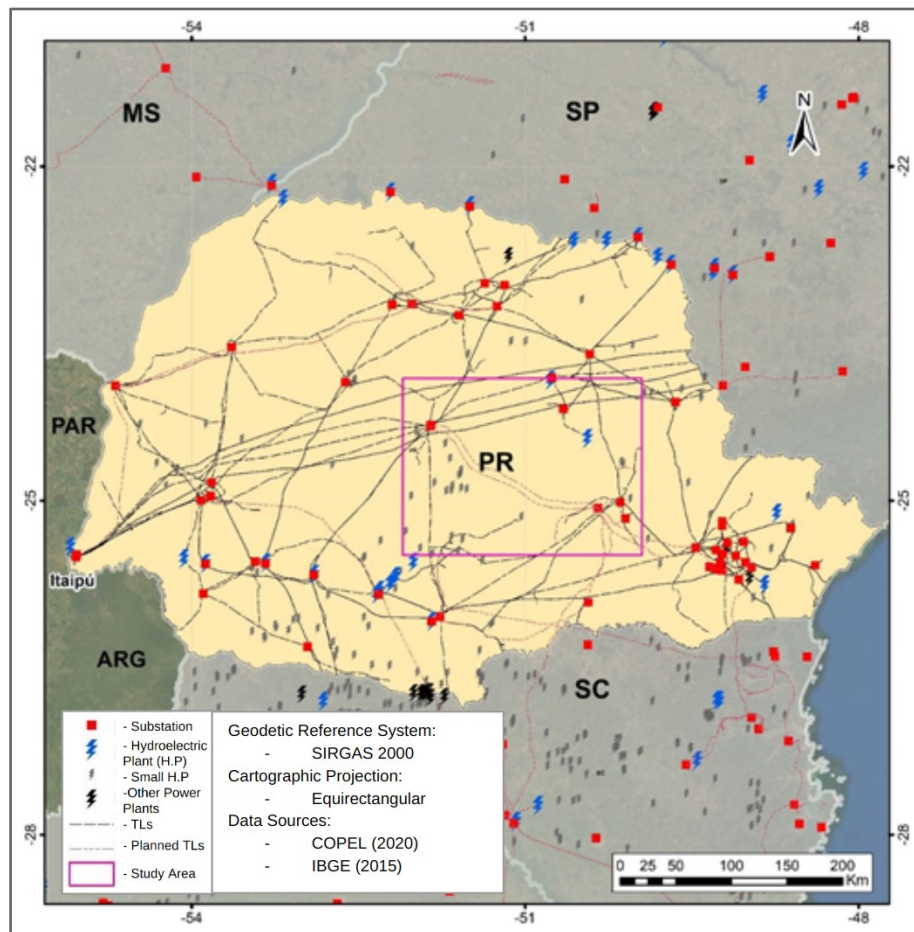
of our method. The article ends in Section 4 with concluding remarks, a discussion of open questions, and suggestions for further work.

METHODOLOGY

As materials for this research, we employed QGIS software and the programming languages Python and R, running in an Intel i7 with 32 GB of ram and a Radeon R7 GPU computer.

The study area, encompassing a portion of the Brazilian state of Paraná is depicted in Figure 01. This area contains 65 municipalities, highlighting *Ponta Grossa and Maringá*, having approximately 37870 km². It was chosen since it is a central area entirely located on Paraná, where there are some planned TLs.

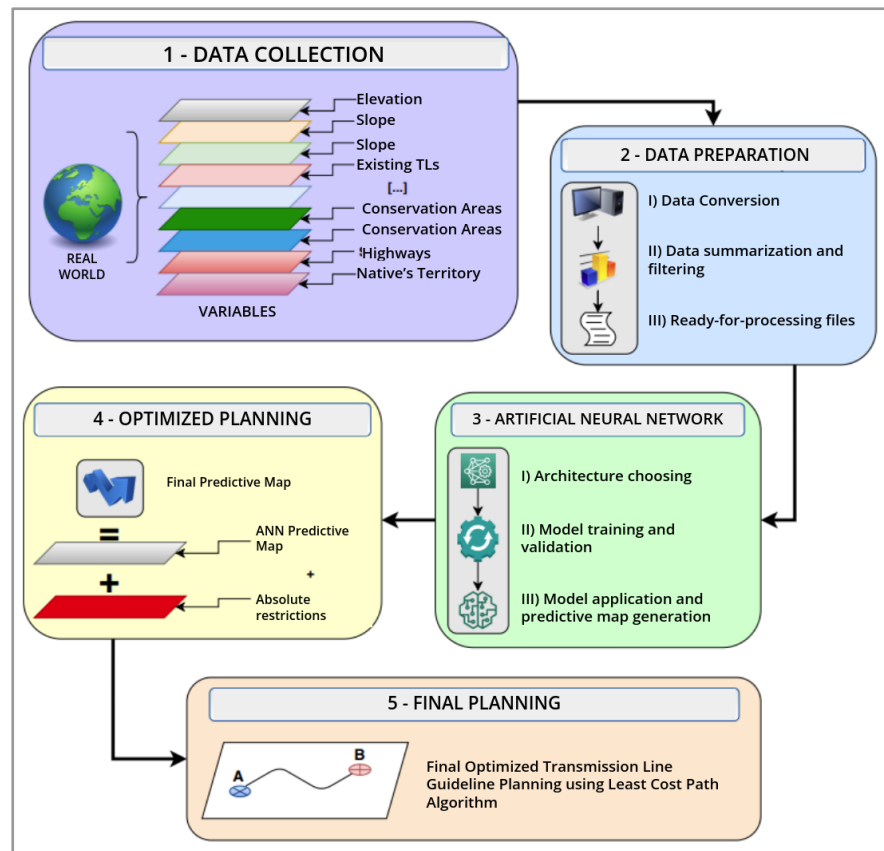
Figure 01: The Study area



Sources: Author (2023).

The employed methodology which is splitted in five steps, is presented at the workflow in figure 02. The five steps are: 1) Data Collection; 2) Data Preparation; 3) Artificial Neural Network; 4) Optimized Planning; 5) Final Planning.

Figure 02: Methodology Workflow



Source: Author (2023).

Regarding the first step, the analysis was made based on data-layers of two types: from both planned and already existing TLs; and thematic reference data that interferes on TLGP. The first group was obtained from COPEL, the TL maintainer on the study area; they have been used to create the positive and negative samples, used both for training and evaluation. The latter group has been collected from open databases, generally public brazilian agencies, referred under parenthesis: 1) hydrography (ANA - *Agência Nacional de Águas*); 2) airports (ANAC - *Agência Nacional de Aviação Civil*); 3) military facilities (BDGEx - *Banco de Dados Geográficos do Exército*); 4) existing TLs (COPEL - *Companhia Paranaense de Energia*); 5) existing highways (DNIT - *Departamento Nacional de Infraestrutura de Transportes*); 6) land use (project MAPBIOMAS); 7) public conservation areas (IAP - *Instituto Ambiental do Paraná*); borders of brazilian municipalities and states (IBGE - *Instituto Brasileiro de Geografia e Estatística*); 8) private conservation areas (INCRA - *Instituto Nacional de Colonização e Reforma Agrária*); 9) native people territory (INDE - *Infraestrutura Nacional de Dados Espaciais*); 10) slope data (INPE); 11) elevation data (INPE - *Instituto Nacional de Pesquisas Espaciais*). This particular set of data layers was chosen based on legal requirements established by the Normative Resolution No. 919, of February 23, 2021 (ANEEL), it's important to notice that for many old TLs there were a different level of restriction, if any.

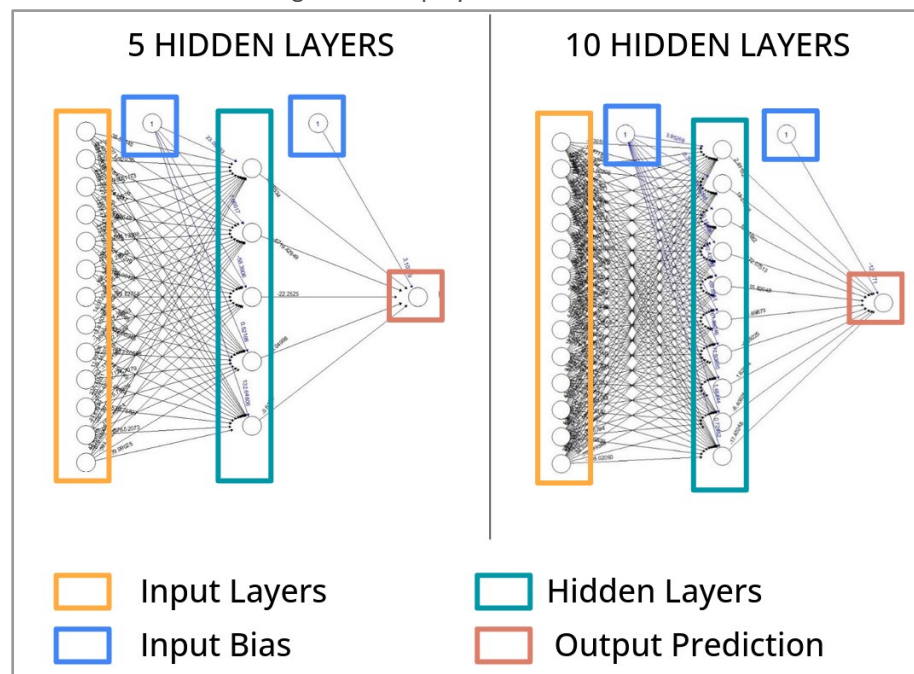
After data acquisition, the next step was data normalization, regarding the gathered data has come in different sizes, resolutions, and sources. We decided that all data should be rasterized and sampled at 30m (as a buffer, except for data source number 2, where we considered 2000 m) by using the arithmetic mean,

with values ranging from 0 to 1, all were projected to an equirectangular projection referred to the WGS84 geodetic reference system. That size isn't arbitrary, it corresponds to the usual width of the environment domain of a transmission line. The categorical data, as data source number 6 was translated into boolean incidence values (a.k.a "one-hot encoding").

After data normalization, we split up TL data into train (60%) and validation (40%) data, by randomly sampling the resulting rasters, resulting in two corresponding .csv files, both with positive and negative samples, which are respectively points with and without Tls. We also considered two scenarios: with 2000 samples and with 20000 samples.

Then we proceeded to the training of the NN. The chosen Neural Network, also called *Neuralnet* (GUNTHER & FRITSCH, 2010) has the following characteristics: it's a Multilayer Perceptron, a classical architecture; the weights and biases were randomly initialized; the minimization was done using the cross-entropy metric; the backpropagation uses a resilient approach (GUNTHER & FRITSCH, 2010) that includes different learning rates for each weight on the network; the weights are initialized randomly using a standard normal distribution; we tested using 5 or 10 Hidden Layers (HL) focusing on the underfitting and overfitting. Then we have four training scenarios: A1 - 5 layers and 2000 samples; A2 - 5 layers and 20000; B1 - 10 layers and 2000 samples; B2: 10 layers and 20000 samples. The stopping criteria, empirically set, was if 10^8 iterations were achieved or if the average prediction error was below 10^{-3} units. A depiction of the employed ANN structure is presented in figure 03.

Figure 03: Employed ANN Structure



Source: Author (2023).

After the training process, we evaluated the model by doing a significance analysis on the values of the variables considered in the process by a sensitivity analysis, namely Lek's Method (GEVREY et. al. 2003), where the "presence" of the variable is progressively increased, to assess each variable's explainability inside

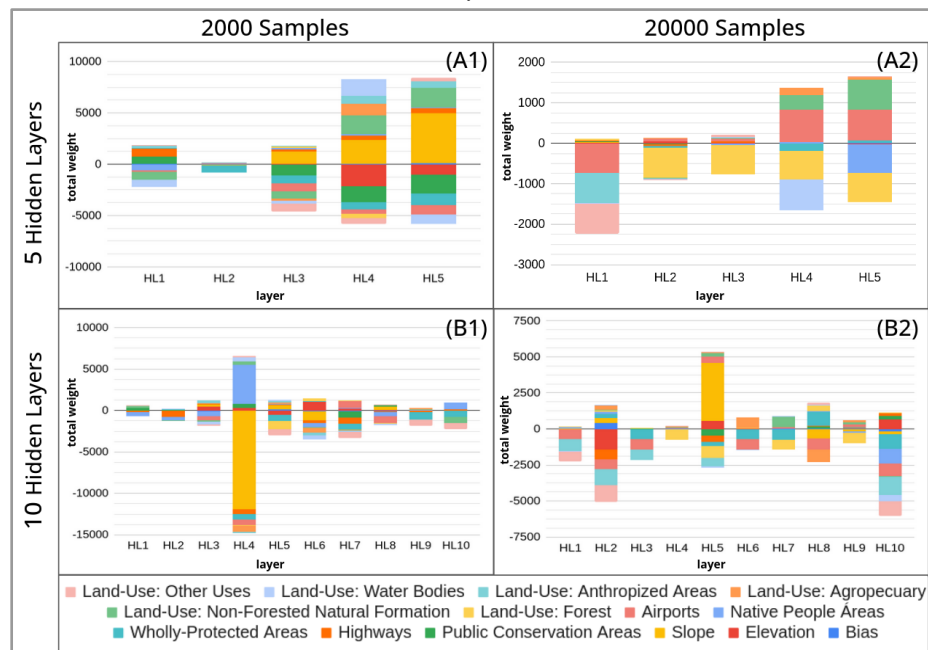
the model. We also checked the confusion matrix, which summarizes the hits and errors among the samples, along with the following metrics (FAWCETT, 2006): Positive Predictive Value (PPV), which is the rate between True Positives (TP) and samples labeled as positives; True Positive Rate (TPP), which is the rate between TP and the samples that are positives; the F1 score that is the harmonic mean between PPV and TPP; and Hit Rate which is the rate of True samples among all the samples. All of them range from 0 (worst) to 1 (best).

As a final result of this research, we have a sample resultant modeling, in the form of a prediction raster map for the whole study area alongside a simulated TLGP, generated using the algorithm LCP - Least-Cost Path (LINDGREN & ERNESTO, 1967) which creates a path between a start and an end point minimizing the sum of weights. The weight of each cell at the prediction map is computed by inputting the incidences of each considered phenomenon into each input layer of the final ANN. We also carried out a relevance comparison of 10 against 5 HL in the RNA structure with the computation of weights; and a comparative analysis of the significance of variables.

RESULTS AND ANALYSIS

As stated in the previous chapter, we have tested four scenarios, combining the number of HL and samples. The main goal of this paper is the achievement of a process for automatic weighting, so we analyze the very weights produced by the different scenarios, as depicted in Figure 04.

Figure 04: the four tested training scenarios and the respective weights for each data layer

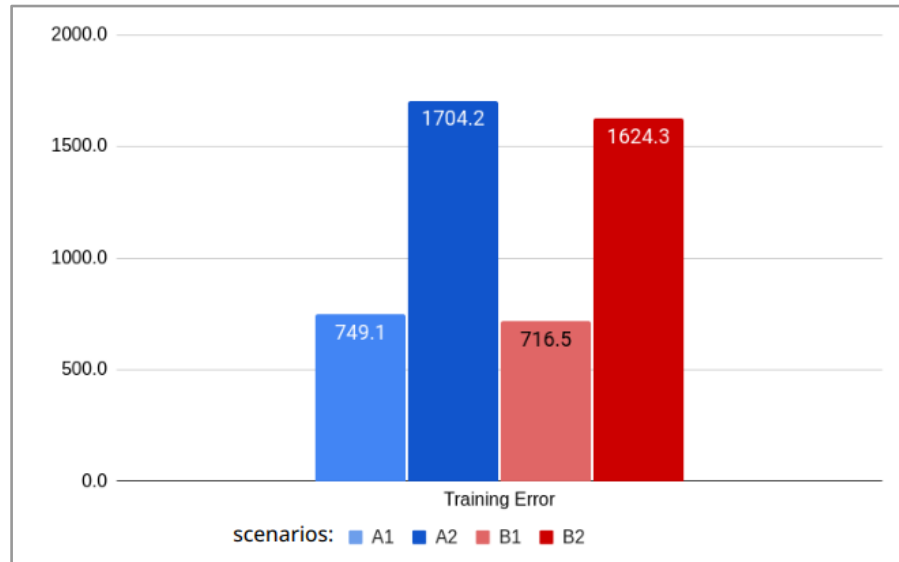


Source: Author (2023).

In Figure 4, we can see that along the ANN there is a complex change of weights among the different layers, reflecting an intricate relation among data towards the decision process, with an apparent predominance of slope in the

weights; in the scenario, A1 the weights are more spread among the classes and at A2 more concentrated among few classes in each layer; in B2 there is a huge predominance at layer number four, awhile on B2 there's a bigger spread. In Figure 5, the training errors are drawn.

Figure 5: the four tested training scenarios and the respective weights for each data layer



Sources: Author (2023).

One can observe that in the scenarios with more samples, there are kind of more stable weights. Figure 5 shows that 5 more layers do not significantly improve the training error (-4.35% from A1 to A2 and -4.68% from B1 to B2), so there's no justification for all the extra computational costs. Then in the forthcoming analysis, only the scenarios with 5 layers (A1 and A2) will be taken into consideration.

Then we looked at the error metrics, analyzing the confusion matrix of the two remaining scenarios, presented in Table 1 and Table 2, respectively.

Table 1: Confusion Matrix and Error Metrics for scenario A1

number of eval. samples:	1200	RESPONSE		True Positive Rate	
		TRUE	FALSE		
SAMPLES	POSITIVE	323	271	0.435	
	NEGATIVE	187	419		
Positive predictive Value		0.544		Hit Rate	0.618
				F1 SCORE	0.484

Source: Author (2023).

Table 2: Confusion Matrix and Error Metrics for scenario A2

number of eval. samples:	12000	RESPONSE		True Positive Rate	
		TRUE	FALSE		
SAMPLES	POSITIVE	3866	2176	0.469	
	NEGATIVE	1573	4385		
Positive predictive Value		0.640		Hit Rate	0.688
				F1 SCORE	0.541

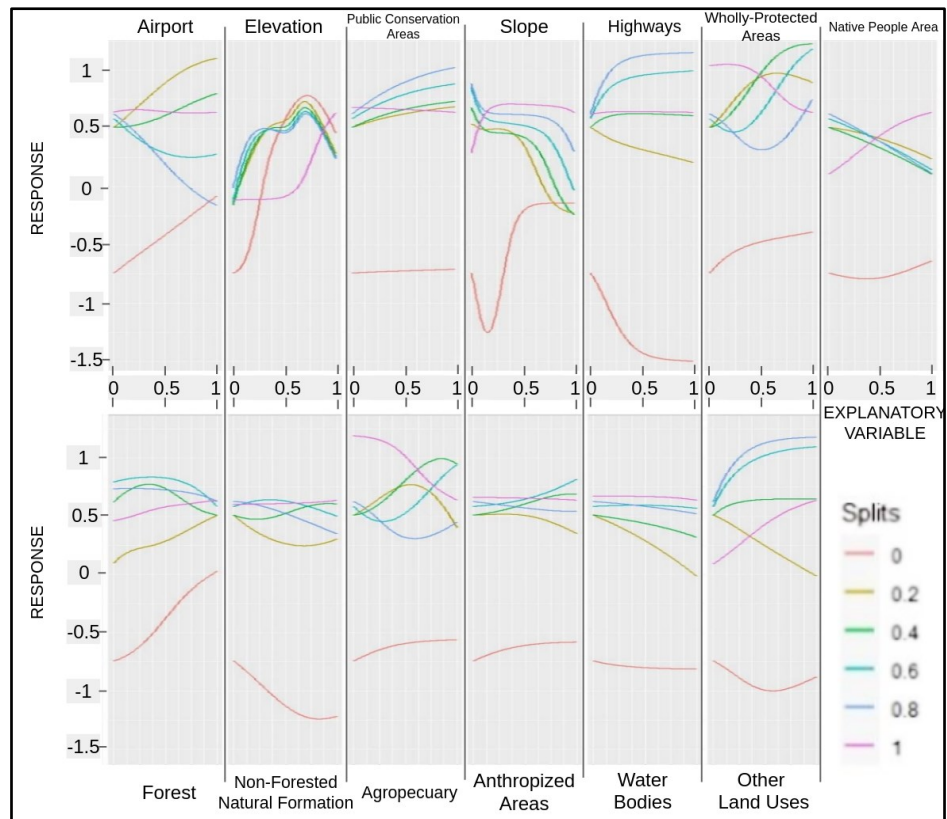
Source: Author (2023).

In Tables 1 and 2 one may see that there is a slight improvement on Hit rate with more samples, but the major improvement are in the PPV, which means that more samples have lead to a minor percent of False Positives i.e a major adherence to the dataset without meaning overfitting, as corroborated by the larger training errors as shown in Figure 5, moreover as the improvements in all error metrics, comparing Tables 1 and 2, it's shown that the larger training errors are not an issue at all. However, in both cases, the F1 score, wich is a good summarizer of the overall quality (FAWCETT, 2006), shows that theres only an intermediate adherence between the predictions and the positive samples. This result was probably due to the fact that many TLs, mainly the older ones, were built without taking many of the analyzed phenomena into consideration and probably, back to the time, there weren't even the tools to carry out those assessments.

The sensitivity analysis (Lek's Method), which was run only for scenario A2 is presented in Figure 6.

Figure 6 shows in most cases non-linear relations between the variables, meaning complex and intricate relations among them. The quantile values at which the remaining explanatory variables were held constant are represented by splits. One can notice that only a line parallel to the X axis would mean that in a particular split that variable means no influence on the model, something that is only near happening at the split 0 of both the input layers "Public Conservation Areas" and "Water Bodies", top splits (0.6, 0.8 and 1) of "Water Bodies" and "Anthropized Areas" and split 1 of "Highways".

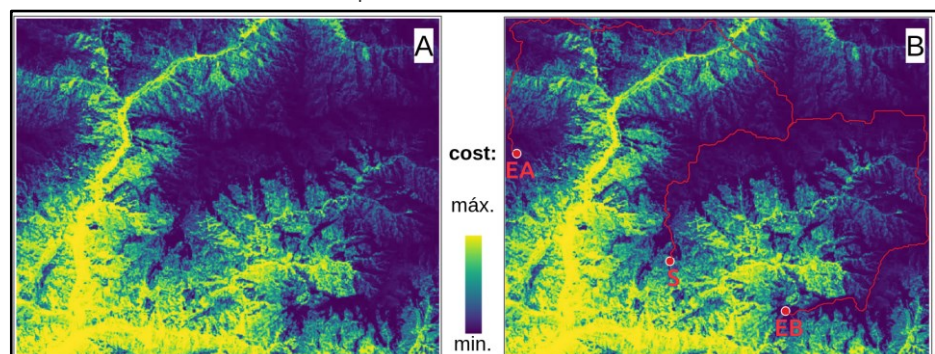
Figure 6: sensitivity analysis



Source: Author (2023).

The final result of this study is a raster map of an LT cost per cell, generated by the final ANN (scenario A2) automated ponderation applied for the whole study area. This prediction map can be used to plan one or multiple optimized TL. In Figure 7, the sample raster map is presented alongside a simulation of two optimized TLGPs using the LCP method.

Figure 7-A: The Sample Generated Raster Map for TL planning at the study area. Figure 7-B: A simulation of optimized TLGP from a fictitious starting point “S” to fictitious ending points “EA” and “EB”.



Source: Author (2023).

The interpretation of the map is that the darker the color is, the more affordable it is to build a TL there. Looking at Figure 7-B one can see that the chosen paths deviate many times the length of a straight line between the start and ending points since there is a big concentration of high-cost raster cells there.

CONCLUSIONS AND FUTURE WORK

In this paper, we proposed the use of AI concepts through artificial neural networks in the problem of transmission line guidelines. This approach provides automation of the process of assigning subjective weights for constraint layers.

Although it is not possible to establish a unit value for the weights of the input variables concerning the predictive variable, the methodology has advantages over the traditional one. This advantage was due to the modeling of the internal relationships between the training covariates themselves, in a way that could be more robust, compared to the standard methods.

The obtained results showed that the use of 5 hidden neurons is enough for this type of trained ANN and did not significantly alter the estimation error metrics nor reflected in the prediction success rate.

Regarding the contribution of the layers, the individual explanatory behavior reinforced the results found in the work of Campos (2014). In this work, the author used only the slope, land use, and distance to highways, existing LT, and the main rivers and waterways. This shows that the currently existing LTs were built with these factors having the most decisive weight. In other words, the economic factor has always been the top priority, to the detriment of social, cultural, historical, and even environmental factors.

Looking at the statistical results of the study, it is clear that at the time of designing the existing LT layout studies, many layers tested here were not taken into consideration decades ago. So, to improve the success of an ANN for path prediction, it is necessary to employ non-biased samples.

In this regard, a secondary focus of this work ends up being to minimize the social, environmental, and economic impacts of this kind of undertaking, since the created model takes into consideration all of these factors.

No input covariates were removed from the study because the intent is that the algorithm shall be an expert system for the layout of new power transmission line studies. So, removing some less explanatory variables could generate a small gain in assertiveness, but could make the algorithm less adaptive to different scenarios as well as would keep neglecting many relevant phenomena.

The improvement of a system like the presented one will provide the creation of a specialist system for analyses of this nature, with several advantages to be exploited. This will provide financial and time savings, besides, contributing to the creation of more sustainable and responsible works in the social sense with preservation of cultural diversity.

So, for system improvement we suggest: investigate the use of a set of simulated and optimal data as input, to minimize the use of biased data in the training phase of the network; test other ANN algorithms; make use of more robust mathematical sampling methods to minimize the divergence effect of the samples; use more accurate or higher resolution data as input parameters, as many data comes with it errors given by interpolation, that can directly affect neural network estimates; manipulating the algorithm to optimize and increase RAM management during the training phase and generation of the predictive surface to allow the use of a bigger amount of samples and to allow the use of higher and more detailed rasters; translate the algorithm to another programming language that allows

better management of the training hardware or perform the translation for training in cloud environments, where the servers have more robust and better configurations than the computer used for the development of this research; develop a metric for improving the explanation of key variables for the analysis; and test the line-sampling approach.

Proposta de ponderação automática na otimização de diretriz de traçado de linhas de transmissão de energia elétrica usando redes neurais artificiais

RESUMO

A ponderação de variáveis na análise multicritério é crucial no planejamento de linhas de transmissão de energia. Este estudo propõe o uso de uma rede neural artificial (RNA) chamada Neuralnet para atribuir automaticamente pesos às variáveis. As entradas para treinamento incluíram fatores sociais, ambientais, culturais e econômicos, como uso do solo, inclinação e áreas indígenas. A RNA foi treinada usando 2 e 20 mil amostras e utilizou a entropia cruzada como métrica de erro. Os resultados mostraram que 5 neurônios ocultos foram suficientes para gerar pesos com uma taxa máxima de sucesso de 69% para a amostra maior. No entanto, essa taxa de sucesso pode ser devido à baixa qualidade das amostras usadas, já que muitas camadas não foram consideradas na construção das linhas de transmissão usadas como amostras positivas. Em geral, o estudo demonstra o potencial de uso da RNA para automatizar a ponderação de variáveis no planejamento de linhas de transmissão de energia, o que pode levar a um refinamento do processo de tomada de decisão.

PALAVRAS-CHAVE: Planejamento de linhas de transmissão. Inteligência Artificial. Geoprocessamento

ACKNOWLEDGEMENTS

To COPEL S.A. for the funding. This research is part of the Project PD-06491-0353/2014, from the COPEL P&D Program. To the Open Data and Open Source communities.

REFERENCES

BRAZILIAN NATIONAL AGENCY OF ELECTRIC ENERGY (Brazil). Planejamento Estratégico ANEEL - Ciclo 2018-2021. Official Report. 2018. Available at: <http://www.aneel.gov.br/documents/656877/2205860/Planejamento+Estrat%C3%A9gico+ANEEL+Ciclo+2018-2021.pdf/16e98df5-5e7d-4b53-8a14-9a9cc9d29433>. Accessed: 22 apr 2022.

BRAZILIAN NATIONAL AGENCY OF ELECTRIC ENERGY (Brazil). Normative Resolution No. 919, of February 23. 2021. Available at: <http://www2.aneel.gov.br/cedoc/ren2021919.pdf>. Accessed: 22 Apr 2022.

BRANS, J.; VINCKE, Ph. A Preference Ranking Organization Method: The PROMETHEE Method for Multiple Criteria Decision-Making. **Management science**, v. 31, n. 6, p. 647-656, 1985.

CAMPOS, P. B. R. **Modelagem de corredores ótimos para linhas aéreas de transporte de energia elétrica**. Master 's Thesis. Graduate Program in Analysis and Modeling of Environmental Systems. Federal University of Minas Gerais. Belo Horizonte, MG, Brazil, 2014. 50 p.

FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861-874, 2006. *ROC Analysis in Pattern Recognition*. ISSN 0167-8655. DOI: 10.1016/j.patrec.2005.10.010. Available at: <https://www.sciencedirect.com/science/article/pii/S016786550500303X>. Accessed: 22 Apr 2022.

FRENCH, S; HOUSTON, G; JOHNSON, C; GLASGOW, J. EPRI-GTC tailored collaboration project: a standardized methodology for siting overhead electric transmission lines. *Environment Concerns in Rights-of-Way Management 8th International Symposium*. p. 221-235, 2008.

GEVREY, M.; DIMOPOULOS, I.; LEK, S. Review and comparison of methods to study the contribution of variables in artificial neural network models. **Ecological modelling**, 160(3), 249-264. 2003.

GEORGESCU, Gh; GAVRILAS, M; CARTINA, G. Aspects concerning power distribution networks planning using artificial intelligence. In: 14th International

Conference and Exhibition on Electricity Distribution. Part 1. Contributions (IEE Conf. Publ. No. 438), v. 6, p. 7-1, 1997.

GÜNTHER, F; FRITSCH, S. neuralnet: Training of Neural Networks. **The R Journal**, v. 2, n. 1, p. 30-38, 2010. DOI: 10.32614/RJ-2010-006. Available at: <https://doi.org/10.32614/RJ-2010-006>. Accessed: 22 Apr 2022.

HEMMATI, R; HOOSMAND, R; KHODABAKHSHIAN, A. Comprehensive review of generation and transmission expansion planning. Wiley Online Library. **IET Generation, Transmission & Distribution**, v. 7, n. 9, p. 955-964, 2013.

KAMYAB, H. R.; ALIPOUR, S. Application of Artificial Neural Network in Landscape Change Process in Gharesou Watershed, Golestan Province. **Iranian Journal of Applied Ecology**, v. 9, n. 2, p. 73-87, 2020.

LAMADRID, A. J.; MANEEVITJIT, S; MOUNT, T. D. The economic value of transmission lines and the implications for planning models. **Energy Economics**, v. 57, p. 1-15, 2016. ISSN 0140-9883. DOI: 10.1016/j.eneco.2016.03.015.

LINDGREN, C. Geography and the properties of surfaces. Proposed solution for the minimum path problem. Technical Report. Harvard University, Cambridge, MA, Lab for Computer Graphics. 1967.

MUGGLETON, S. Alan Turing and the development of Artificial Intelligence. **AI communications**, v. 27, n. 1, p. 3-10, 2014.

PECHANEC, V.; BURIAN, J.; DOBEŠOVÁ, Z. Integration Neural Networking and GIS in Modeling Landscape Changes. **International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management**, SGEM, v. 1, n. 2, p. 651-658, 2014.

PERRAS, S. Electricity transmission line planning: Success factors for transmission system operators to reduce public opposition. Technische Universität Dresden, 2015. Available at: <https://core.ac.uk/download/pdf/236372191.pdf>. Accessed: 22 Apr 2022.

ROY, B. The outranking approach and the foundations of ELECTRE methods. **Theory and Decision**, v. 31, n. 1, p. 49-73, 1991. ISSN 1573-7187.. Available at: <https://doi.org/10.1007/BF00134132>. Accessed: 22 Apr 2022.

SAATY, R.W. The analytic hierarchy process—what it is and how it is used. **Mathematical Modelling**, v. 9, n. 3, p. 161-176, 1987. ISSN 0270-0255. DOI: 10.1016/0270-0255(87)90473-8. Available at:

<https://www.sciencedirect.com/science/article/pii/0270025587904738>.

Accessed: 22 Apr 2022.

SPATH, L.; SCOLOBIG, A. Stakeholder empowerment through participatory planning practices: The case of electricity transmission lines in France and Norway. **Energy Research & Social Science**, v. 23, p. 189-198, 2017. ISSN 2214-6296. DOI: 10.1016/j.erss.2016.10.002. Available at:

<https://www.sciencedirect.com/science/article/pii/S2214629616302377>.

Accessed: 22 Apr 2022.

YANG, D.; HE, T.; DU, B.; WANG, S.; ZHANG, Z.; ZHANG, S. Transmission Line Planning Based on Artificial Intelligence in Smart Cities. **Mobile Information Systems**, vol. 2022, p. 1-8, 31 Mar. 2022. Available at: <https://doi.org/10.1155/2022/2010189>. Accessed: 17 May 2023.

Recebido: 08 mar. 2023

Aprovado: 12 jun. 2023

DOI: 10.3895/rbgeo.v12n1.16510

Como citar: BOSSOLA, L. H.; SANTOS, D. R.; VESTENA, K. M.; ANTONIO, N. D.. Proposta de ponderação automática na otimização de diretriz de traçado de linhas de transmissão de energia elétrica usando redes neurais artificiais. **R. bras. Geom.**, Curitiba, v. 12, n. 1, p. 038-052, jan./jun. 2024. Disponível em: <<https://periodicos.utfpr.edu.br/rbgeo>>. Acesso em: XXX.

Correspondência:

Luiz Henrique Bossola

Rua Desembargador Otavio do Amaral, 716 ap 15, CEP 80730-400, Curitiba, Paraná, Brasil.

Direito autoral: Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

