This is the Accepted Manuscript for "**Investigation of the influence of nonoccurrence sampling on Landslide Susceptibility Assessment using Artificial Neural Networks**", published on Catena in March 2021.

Link to the formal publication:

doi: https://doi.org/10.1016/j.catena.2020.105067

in accordance with Elsevier Sharing Policy (https://www.elsevier.com/about/policies/sharing).

e. **"Original Author"** means, in the case of a literary or artistic work, the individual, individuals, entity or entities who created the Work or if no individual or entity can be identified, the publisher; and in addition (i) in the case of a performance the actors, singers, musicians, dancers, and other persons who act, sing, deliver, declaim, play in, interpret or otherwise perform literary or artistic works or expressions of folklore; (ii) in the case of a phonogram the producer being the person or legal entity who first fixes the sounds of a performance or other sounds; and, (iii) in the case of broadcasts, the organization that transmits the broadcast.

f. **"Work"** means the literary and/or artistic work offered under the terms of this License including without limitation any production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression including digital form, such as a book, pamphlet and other writing; a lecture, address, sermon or other work of the same nature; a dramatic or dramatico-musical work; a choreographic work or entertainment in dumb show; a musical composition with or without words; a cinematographic work to which are assimilated works expressed by a process analogous to cinematography; a work of drawing, painting, architecture, sculpture, engraving or lithography; a photographic work to which are assimilated works expressed by a process analogous to photography; a work of applied art; an illustration, map, plan, sketch or three-dimensional work relative to geography, topography, architecture or science; a performance; a broadcast; a phonogram; a compilation of data to the extent it is protected as a copyrightable work; or a work performed by a variety or circus performer to the extent it is not otherwise considered a literary or artistic work.

g. **"You"** means an individual or entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation.

h. **"Publicly Perform"** means to perform public recitations of the Work and to communicate to the public those public recitations, by any means or process, including by wire or wireless means or public digital performances; to make available to the public Works in such a way that members of the public may access these Works from a place and at a place individually chosen by them; to perform the Work to the public by any means or process and the communication to the public of the performances of the Work, including by public digital performance; to broadcast and rebroadcast the Work by any means including signs, sounds or images.

i. **"Reproduce"** means to make copies of the Work by any means including without limitation by sound or visual recordings and the right of fixation and reproducing fixations of the Work, including storage of a protected performance or phonogram in digital form or other electronic medium.

2. Fair Dealing Rights. Nothing in this License is intended to reduce, limit, or restrict any uses free from copyright or rights arising from limitations or exceptions that are provided for in connection with the copyright protection under copyright law or other applicable laws.

3. License Grant. Subject to the terms and conditions of this License, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) license to exercise the rights in the Work as stated below:

a. to Reproduce the Work, to incorporate the Work into one or more Collections, and to Reproduce the Work as incorporated in the Collections; and,

b. to Distribute and Publicly Perform the Work including as incorporated in Collections.

The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats, but otherwise you have no rights to make Adaptations. Subject to 8(f), all rights not expressly granted by Licensor are hereby reserved, including but not limited to the rights set forth in Section 4(d).

4. Restrictions. The license granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a. You may Distribute or Publicly Perform the Work only under the terms of this License. You must include a copy of, or the Uniform Resource Identifier (URI) for, this License with every copy of the Work You Distribute or Publicly Perform. You may not offer or impose any terms on the Work that restrict the terms of this License or the ability of the recipient of the Work to exercise the rights granted to that recipient under the terms of the License. You may not sublicense the Work. You must keep intact all notices that refer to this License and to the disclaimer of warranties with every copy of the Work You Distribute or Publicly Perform. When You Distribute or Publicly Perform the Work, You may not impose any effective technological measures on the Work that restrict the ability of a recipient of the Work from You to exercise the rights granted to that recipient under the terms of the License. This Section 4(a) applies to the Work as incorporated in a Collection, but this does not require the Collection apart from the Work itself to be made subject to the terms of this License. If You create a Collection, upon notice from any Licensor You must, to the extent practicable, remove from the Collection any credit as required by Section 4(c), as requested.

b. You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital file-sharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

c. If You Distribute, or Publicly Perform the Work or Collections, You must, unless a request has been made pursuant to Section 4(a), keep intact all copyright notices for the Work and provide, reasonable to the medium or means You are utilizing: (i) the name of the Original Author (or pseudonym, if applicable) if supplied, and/or if the Original Author and/or Licensor designate another party or parties (e.g., a sponsor institute, publishing entity, journal) for attribution ("Attribution Parties") in Licensor's copyright notice, terms of service or by other reasonable means, the name of such party or parties; (ii) the title of the Work if supplied; (iii) to the extent reasonably practicable, the URI, if any, that Licensor specifies to be associated with the Work, unless such URI does not refer to the copyright notice or licensing information for the Work. The credit required by this Section 4(c) may be implemented in any reasonable manner; provided, however, that in the case of a Collection, at a minimum such credit will appear, if a credit for all contributing authors of Collection appears, then as part of these credits and in a manner at least as prominent as the credits for the other contributing authors. For the avoidance of doubt, You may only use the credit required by this Section for the purpose of attribution in the manner set out above and, by exercising Your rights under this License, You may not implicitly or explicitly assert or imply any connection with, sponsorship or endorsement by the Original Author, Licensor and/or Attribution Parties, as appropriate, of You or Your use of the Work, without the separate, express prior written permission of the Original Author, Licensor and/or Attribution Parties.

d. For the avoidance of doubt:
    i. **Non-waivable Compulsory License Schemes**. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme cannot be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License;

    ii. **Waivable Compulsory License Schemes**. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme can be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License if Your exercise of such rights is for a purpose or use which is otherwise than noncommercial as permitted under Section 4(b) and otherwise waives the right to collect royalties through any statutory or compulsory licensing scheme; and,

iii. **Voluntary License Schemes**. The Licensor reserves the right to collect royalties, whether individually or, in the event that the Licensor is a member of a collecting society that administers voluntary licensing schemes, via that society, from any exercise by You of the rights granted under this License that is for a purpose or use which is otherwise than noncommercial as permitted under Section 4(b).

e. Except as otherwise agreed in writing by the Licensor or as may be otherwise permitted by applicable law, if You Reproduce, Distribute or Publicly Perform the Work either by itself or as part of any Collections, You must not distort, mutilate, modify or take other derogatory action in relation to the Work which would be prejudicial to the Original Author's honor or reputation.

5. Representations, Warranties and Disclaimer

UNLESS OTHERWISE MUTUALLY AGREED BY THE PARTIES IN WRITING, LICENSOR OFFERS THE WORK AS-IS AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE WORK, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTIBILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT, OR THE ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OF ABSENCE OF ERRORS, WHETHER OR NOT DISCOVERABLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO SUCH EXCLUSION MAY NOT APPLY TO YOU.

6. Limitation on Liability. EXCEPT TO THE EXTENT REQUIRED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY FOR ANY SPECIAL, INCIDENTAL, CONSEQUENTIAL, PUNITIVE OR EXEMPLARY DAMAGES ARISING OUT OF THIS LICENSE OR THE USE OF THE WORK, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

7. Termination

a. This License and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this License. Individuals or entities who have received Collections from You under this License, however, will not have their licenses terminated provided such individuals or entities remain in full compliance with those licenses. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this License.

b. Subject to the above terms and conditions, the license granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different license terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

8. Miscellaneous

a. Each time You Distribute or Publicly Perform the Work or a Collection, the Licensor offers to the recipient a license to the Work on the same terms and conditions as the license granted to You under this License.

b. If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

c. No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

d. This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You.

e. The rights granted under, and the subject matter referenced, in this License were drafted utilizing the terminology of the Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979), the Rome Convention of 1961, the WIPO Copyright Treaty of 1996, the WIPO Performances and Phonograms Treaty of 1996 and the Universal Copyright Convention (as revised on July 24, 1971). These rights and subject matter take effect in the relevant jurisdiction in which the License terms are sought to be enforced according to the corresponding provisions of the implementation of those treaty provisions in the applicable national law. If the standard suite of rights granted under applicable copyright law includes additional rights not granted under this License, such additional rights are deemed to be included in the License; this License is not intended to restrict the license of any rights under applicable law.

Investigation of the influence of nonoccurrence sampling on Landslide Susceptibility Assessment using Artificial Neural Networks

Luísa Vieira Lucchese[1,3], Guilherme Garcia de Oliveira[2], Olavo Correa Pedrollo[1]

Abstract

Landslide susceptibility assessment using Artificial Neural Networks (ANNs) requires occurrence (landslide) and nonoccurrence (not prone to landslide) samples for ANN training. We present empirical evidence that a priori intervention on the nonoccurrence samples can produce models that are improper for generalization. Thirteen nonoccurrence cases based on GIS data from Rolante River basin (828.26 km^2) in Brazil are studied, divided in three groups. The first group was based on six combinations of buffers with different minimum and maximum distances from the mapped scars (BO). The second group (RO) acquired nonoccurrence only from a rectangle in the lowlands, known for not being susceptible to landslides. For BR, six alternatives respectively to BO were presented, with the inclusion of nonoccurrence samples acquired from the same rectangle used for RO. Accuracy (acc) and the Area Under Receiving Operating Characteristic Curve (AUC) were calculated. RO resulted in perfect discrimination between susceptible and not susceptible to landslides (acc=1 e AUC=1). This occurred

[1]Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, 91501-970, Porto Alegre, RS, Brazil
[2]Departamento Interdisciplinar. Universidade Federal do Rio Grande do Sul. Rodovia RS 030, 11700, km 92. Emboaba, Tramandaí/RS, Brazil. CEP 95590-000.
[3]Corresponding author.
E-mail address: luisa.lucchese@ufrgs.br (Luísa Lucchese)
Declarations of interest: none.

because the model simply provided susceptible classification to points in which attributes are different from those in the rectangle, harming the classification of nonoccurrence sampling points outside the rectangle. RO map shows large areas classified as susceptible which are known to be non-susceptible. In BR, sampling points from the rectangle, which are easy to classify, were added to the verification sample of BR. Average acc for BO 00m (minimum buffer distance to scars of 0 m): 89.45%, average acc for BR 00m: 92.33%, average AUC for BO 00m: 0.9409, average AUC for BR 00m: 0.9616. Maps of groups BO and BR were alike. This indicates that metrics can be artificially risen if biased samples are added, although the final map is not visibly affected. To avoid this effect, the employment of easily classifiable samples, generated based on expert knowledge, should be made carefully.

Keywords: landslides, mass movements, South America, Rio Grande do Sul Brazil, sediment transport, geomorphology

1. Introduction

Landslide susceptibility assessment is the process that establishes the likelihood of landslide occurrence in a given area, using suitable terrain factors (Sorriso Valvo, 2002). A possible way of assessing natural disasters hazard is to produce, using Geographic Information Systems (GIS) techniques, maps of susceptibility to the disaster. In order to generate these maps, data from previous landslides is usually necessary. This relates to one of the general principles of landslide hazard zonation, that is the past is the key to the future (Varnes, 1984; Fell et al., 2008).

Landslide susceptibility assessment can be performed by a range of methods, that are comprised of two main approaches: qualitative and quantitative. Qualitative approaches are usually based on on-site observations and combinations of index maps elaborated by experts. One example is Anbalagan (1992) which used the

Landslide Hazard Evaluation factor. Aleotti and Chowdhury (1999) discuss that, decades ago, on-site survey was essentially the only option available that was not prone to subjectivity involved, though it could be costly and dangerous. On the other hand, quantitative approaches are based on statistical models (Lee and Min, 2001; Regmi et al., 2014; Hussin et al., 2016) or geotechnics approaches (Gökceoglu and Aksoy, 1996; Fall et al., 2006; Gutiérrez-Martín, 2020). Artificial Intelligence models may be considered included in the statistical category, within the quantitative approach.

Artificial Intelligence (AI) is the theory and development of computer systems with the ability to act resembling human intelligence. To date, several studies have used AI methods for Landslide Susceptibility Mapping. AI methods used in this knowledge area include but are not limited to Artificial Neural Networks (ANNs) - used in the present paper as well as in the papers of Lee et al. (2004) and Ermini et al. (2005), but also Random Forest (RF) - used by Catani et al. (2013), Pourghasemi and Kerle (2016) and Dou et al. (2019) -, Rotation Forest (RoF) - used by Chen et al. (2017b) -, Fuzzy Inference Systems (FIS) - employed by Ercanoglu and Gokceoglu (2002) and Kanungo et al. (2006) -, Logistic Regression (LR) - employed by Ayalew and Yamagishi (2005) and Lee (2005) -, and Naïve Bayes (NB) - used by Bui et al. (2012).

Some authors performed comparisons between AI methods. Most of the papers that compared ANN to other methods for landslide susceptibility assessment demonstrated that ANN models perform better than their counterparts. Yesilnacar and Topal (2005) compared ANNs to LR and calculated the global accuracy for both, which was 82.1% for ANN and 79.6% for LR. A comparison between Dempster–Shafer, LR and ANN was made by Chen et al. (2017c). The accuracies calculated for the validation set for the three methods were, respectively, 61.39%, 68.94% and 69.92%. Dou et al. (2018) used both ANN and LR, and concluded

ANN provided a higher accuracy. These findings are also supported by Gong et al. (2018), that also compared ANN and LR, obtaining 82.6% and 75.4% accuracies for ANN and LR, respectively. Braun et al. (2019) compared the accuracy of three methods, namely, Decision Trees, Bayesian Networks and ANNs, and ANNs were better at classifying the landslide areas, with some level of false alarm on the classification maps. Overall, many researchers found that ANNs tend to provide good accuracy for landslide prediction.

A large and growing body of literature has investigated the use of ANNs for landslide susceptibility assessment. As ANNs are universal approximators (Hornik et al., 1989), using the right training dataset, number of hidden neurons and activation function for these neurons, one can possibly approximate any existing function. However, it was not until the 2000s that studies on using ANNs for landslide susceptibility assessment were first published. Some pioneer papers on this subject were Lee et al. (2004), which used seven attributes to train the ANN, Ermini et al. (2005), which used a Probabilistic Neural Network, and Gomez and Kavzoglu (2005), which employed nine different terrain attributes for training. More recently, many studies have been conducted in this area, such as the works of Dou et al. (2015), which selected six out of 14 original attributes for ANN training and employed them for landslide susceptibility assessment on Osado Island, Japan, Chen et al. (2017a), that compared three different types of models, including ANNs, and Braun et al. (2019), which used ANN models for the same objective in Honduras.

Sampling locations and techniques used to train, validate and verify the ANNs and maps generated are, at some extent, unexplored in the methodology presentation of published papers in landslide susceptibility assessment. As Zhu et al. (2018) emphasize, the employment of nonoccurrence samples is very important to constrain the over-prediction of high susceptibility. Some papers do not provide any information about nonoccurrence sites. Ortiz and Martínez-Graña (2018) used half of landslide sites for training, but presented no information about nonoccurrence samples used. Xiong et al. (2019) present some information about the attributes in the occurrence samples, but nonoccurrence is not commented. Bui et al. (2016) separates the landslide inventory in landslides used to train and to verify the ANN model, but they do not provide the information about nonoccurrence sites used to train. In the work of Can et al. (2019), it is stated that equal numbers (196) of occurrence and nonoccurrence samples were used to train the ANN, but neither the location nor the parameters of nonoccurrence samples are presented. On the other hand, Zare et al. (2013) presented the frequency ratio between occurrence and nonoccurrence for each factor used to train the ANN, and Pham et al. (2017) showed the location of every point.

Still, to date there has been little agreement on the methods used to acquire nonoccurrence (safe places, not prone to landslides) samples. Pradhan and Lee (2010), Dou et al. (2018), Polykretis and Chalkias (2018) and Shirzadi et al. (2019) used random points outside the landslide scars. Braun et al. (2019) employed all locations with no landslides recorded as nonoccurrence samples, and created copies of landslide-prone samples in order to have a balance of the two classes on the training set. Merghadi et al. (2018) divided the randomly sampled points in ten divisions, using nine of them to train and one to validate, and repeated the sampling procedure five times. Their evaluation metrics were averaged between the 50 models trained. Pham et al. (2017) presented the location of all occurrence and nonoccurrence points used for training and for validation on a map, on which it is possible to see the points are well-distributed. Gomez and Kavzoglu (2005) selected nonoccurrence samples from places where landslide initialization is not likely to happen, such as on river channels and terrains with slope angles lower than five degrees. The accuracy achieved by the ANN for areas mapped as not susceptible to landslides was about 95%. The generated map shows large areas as highly susceptible and appears to had been strongly influenced by Elevation attribute. Choi et al. (2010) executed an ANN trained for one location in another location. Analogously, in their study, nonoccurrence samples were taken from points in which the slope was zero. When the ANN models are applied to the areas for which they were calculated, large areas appear to have susceptibility indexes over 0.8.

Xiao et al. (2010) did not use ANNs but a Generalized Additive Model, which is a variation of LR, to propose a method for artificially generating nonoccurrence samples. This method is based on occurrence samples and is called Target Space Exteriorization Sampling. Samples are created by changing the values of one or more attributes so that they are out of the range considered for susceptibility. They compared their method to other two approaches. One of them is using random samples with a minimum distance of 85 m from the scars (called Buffer Controlling Sampling). The other one is using these random samples with outliers filtered out (called Iteratively Refined Sampling). They concluded that the proposed sampling method provided satisfactory evaluation metrics and a more desirable probability distribution of outputs than the other two. Filtering outliers out generated higher metrics than using the original (random) dataset. Analyzing this finding from a different point of view, higher evaluation metrics could possibly be caused by this intervention, which eliminates some of the nonoccurrence samples that would be hardest to classify.

In the paper of Hong et al. (2019), that compared four methods for nonoccurrence sampling using RF, the geographical locations from which the nonoccurrence samples could be acquired by each sampling method are shown. Also, it is possible to see that methods which exclude wider areas around the landslide scars result in susceptibility maps with large areas classified as susceptible, and this effect intensifies as the balance between occurrence and nonoccurrence samples changes by the removal of some nonoccurrence samples. This was observed by Hong et al. (2019) but the underlying causes were not researched.

Zhu et al. (2019) employed a similarity-based approach for acquiring nonoccurrence samples for landslide susceptibility mapping using SVM, RF and LR. A reliability index was calculated for nonoccurrence samples, based on how much those samples attributes differ from the occurrence samples attributes. At some extent, this can be considered expert knowledge intervention on the dataset. Zhu et al. (2019) acknowledged that using samples with reliability index over 0.5 harms the correct classification of nonoccurrence areas, generating maps that overestimate susceptible areas, even if the models based on these samples showed the highest accuracy metrics. Their explanation is that, in these cases, occurrence and nonoccurrence samples are too different. Nevertheless, the possible non-applicability of metric comparisons for the analyzed cases was not discussed. In theory, higher accuracies indicate better modeling capabilities, however, Zhu et al. (2019) results can be seen in a new light. A possibility to be contemplated is that nonoccurrence samples with higher reliability indexes can be easier to classify by a data-driven model than samples with low reliability indexes. However, this may bias the evaluation metrics calculated based on them and possibly make them unsuited for direct comparisons.

About the occurrence sampling methods, a number of studies have begun to examine sampling strategies for landslide hazard assessment. Süzen and Doyuran (2004) proposed the seed cells method, in which the landslide occurrence samples would be acquired from zones that are considered to represent the undisturbed (i.e. before the occurrence of the event) morphological conditions, acquired from the vicinity of the landslide polygon. The location of these seed cells is acquired using a buffer from the mapped scars, upstream of the scars.

Wang et al. (2013) used logistic regression and employed a 50 m buffer of the landslide seed cells. Only the location of occurrence samples is given by the seed cell method, the location of nonoccurrence samples is not discussed. Yilmaz (2010) investigated three possible techniques to select which area should be included in the occurrence samples, using the location of landslide scars. However, he did not elucidate how the nonoccurrence samples should be taken. Yao et al. (2008), on the other hand, compared the use of two-class (occurrence and nonoccurrence) Supporting Vector Machine (SVM) with one-class (occurrence only) correspondent model, and detected that results improved when using the two-class model, noticeably showing that nonoccurrence samples matter.

Evaluation metrics such as Area Under Receiver Operating Characteristic (ROC) Curve (AUC) (DeLong et al., 1988) and accuracy (acc) are reported in many papers of the research area. Although, in many of these papers, the base samples (and their locations) used for calculating the evaluation metrics are not presented. In some papers, only the test area was used for these calculations (Kawabata and Bandibas, 2009; Merghadi et al., 2018), or a separate statistic is calculated for the test set (Kumar et al., 2018). In others, the whole dataset composed of training, validation - if present - and verification samples is used to calculate the evaluation metrics (García-Rodríguez and Malpica, 2010; Zhu et al., 2018). This is relevant because some groups of samples are remarkably easier to classify than others. Does the criterion used to select nonoccurrence samples for landslide susceptibility mapping influence the generalization ability of the ANN model? Also, is it possible that, when we select easily classifiable samples, we facilitate the modeling process, and attain inflated evaluation metrics, without necessarily this correlating to a better generalization ability?

Notably, the results of the evaluation metrics will not be reliable if the experiment planning is not well conduced regarding the verification samples.

One should notice that the best criteria are the evaluation metrics, but this is only true if the verification sample, used for comparison between models, is representative of the application areas studied, and subjected to same triggering factors. This paper discusses the necessity of classifications that are not obvious, to avoid biased resulting models. We therefore provide empirical evidence that the detection of bias on the resulting models is hard if the nonoccurrence verification samples are acquired from obvious nonoccurrence areas.

Overall, the studies here depicted highlight the need for more investigation into the samples used to train, validate (when applicable) and verify the employed ANNs. In this paper, we aim to show how different nonoccurrence sampling techniques can influence the resulting susceptibility map and statistics. Notably, nonoccurrence sampling based on expert knowledge of susceptibility, here represented by acquiring nonoccurrence samples from the lowlands. If sampling is done exclusively in the scars proximity, there is a risk that the model, because it does not mark the lowlands as occurrence sites, fails when applied to a larger area. This would make the specialist to limit the model application area to places where, according to their expert knowledge, have greater probability to be susceptible. In this case, they would determine (therefore, intervening) that the other regions would be nonoccurrence sites. The inclusion of a rectangle in the lowlands is based on making the model identify areas that are notably not susceptible to landslides without the need to intervene during the application of the model, intervening, although, in the sampling process. The idea is interesting, but it should be analyzed so that we provide evidence on whether it is successful.

Through a rather empirical approach, we aim to investigate how a sampling procedure in which the nonoccurrence points are acquired from areas known not to be prone to landslides influences the evaluation metrics. Thus, metrics can be affected in an apparently positive way, although inaccurate susceptibility maps can be generated. This sampling procedure, portrayed here, is based on the acquisition of samples from a lowland area on the valley, known not to be prone to landslides. In opposition, samples can be acquired from close locations to the places where the landslides occurred. This possibility is also investigated here.

2. Study Area

Our study is applied to the area of Rolante River basin, located in Rio Grande do Sul state, in Brazil. In Fig. 1 the location of the area is indicated. Zooming into it, we observe the rectangular studied area. Rolante River basin is highlighted for reference. It is a watershed with high sloped scarps, located in a predominantly rural area. In this figure, terrain elevation is plotted on background, for reference. The scars of the 2017 landslide events are also plotted.

On January 5th, 2017, a series of landslides occurred in the area (scars marked on Fig. 1). They were triggered by intense rainfall on that day and showers that had been occurring for days. The exact amount of rain in this area remains unknown. It was a convective rainfall, very localized, typical of summer season, therefore hard to be correctly measured by satellites. According to Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) (Huffman et al., 2007), a spatial average of 142.44 mm of rainfall was recorded in the studied area, during two and a half days (57h). In opposition, a technical report released (Secretaria Estadual do Meio Ambiente and Grupo de Pesquisa em Desastres Naturais, 2017) provided rainfall measurements of seven local farmers in their properties, varying from 90 to 272 mm in 24 h, although caution should be exercised since some of their equipment did not

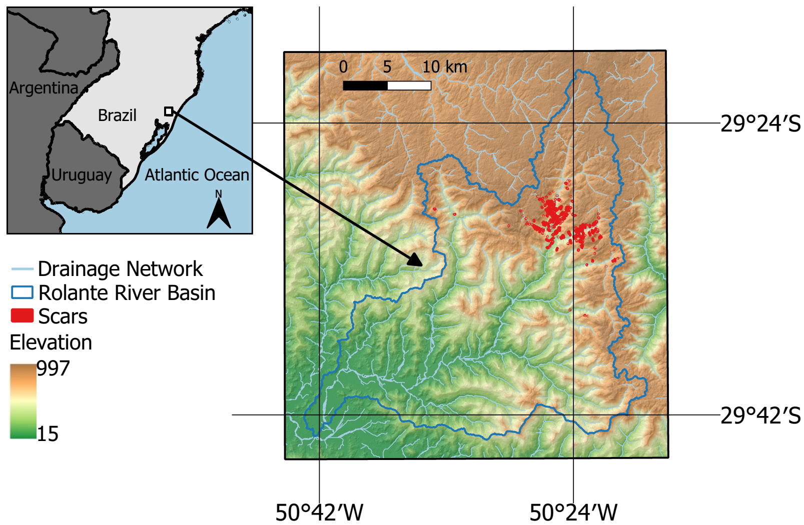

Figure 1: Map of the study area and the location of the landslide scars.

follow technical specifications.

Even though no one was physically hurt by the natural hazard, the need for zonation in this area is a pressing issue. Zonation starts with a hazard susceptibility assessment, in this case, a landslide susceptibility assessment.

The original data for this area is from Advanced Land Observing Satellite-1 (ALOS), a mission of the Japan Aerospace Exploration Agency (JAXA) that used the Phased Array type L-band Synthetic Aperture Radar (PALSAR) to map the surface of some areas on Earth (ASF DAAC, 2015). The DEM was downloaded through Alaska Satellite Facility (ASF) Distributed Active Archive Center (DAAC), that operates the North American Space Agency (NASA) archive of Synthetic Aperture Radar (SAR) data from a variety of satellites and aircraft, in support of NASA's Earth Science Data and Information System (ESDIS) project. According to Arnone et al. (2016), the DEM with 10 m spatial resolution was better suited for landslide susceptibility mapping using

ANNs than the ones with 2, 5, 20, 30, 40, and 50 m resolution. Among the DEMs available for this region, the one with the highest spatial resolution is the one that was chosen by the present authors for this research, which is ALOS PALSAR, with 12.5 m resolution. The DEM and the scar inventory are the only two data sources used for this model, every other attribute is generated based on the DEM by using software QGis. Ten attributes were generated, plus the terrain elevation, which is the DEM itself. The generated attributes are (Fig. 2):

- Aspect: the orientation, in degrees, of the hill slope;
- Hillshade: a grayscale 3D representation of the surface;
- Natural Logarithm of Flow Accumulation: flow accumulation based on flow direction, in log scale;
- Planar Curvature: the horizontal curvature of terrain;
- Profile Curvature: the vertical curvature of terrain;
- Slope: the declivity of the slope, in degrees;
- Slope Length and Steepness Factor: also known as LS-factor, it is a term of Universal Soil Loss Equation (USLE) for soil erosion;
- Topographic Wetness Index (TWI): an index that considers uphill drainage area and slope, $ln(Area/tan(SLOPE))$;
- Valley Depth: the vertical distance to the channel network base level;
- Vertical Distance to Channel Network (VDCN): the vertical distance to the nearest draining channel.

Fig. 2 shows the spatial distribution of the attributes. The attributes chosen for this application are believed to provide significant data to the model, in

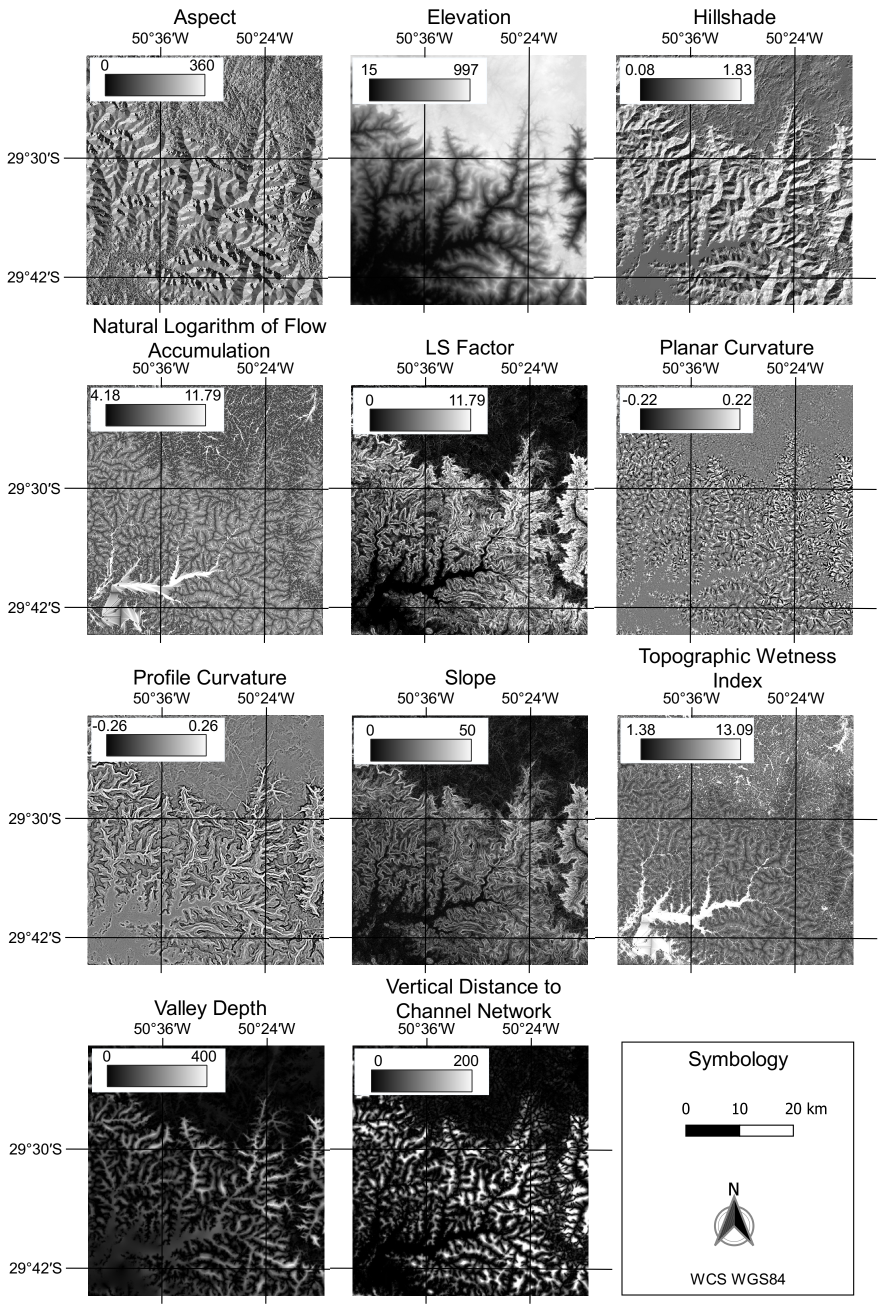

Figure 2: Attributes on the study area.

terms of relationships between pixels, and to provide information about the relative location of each pixel on the basin. Elevation can provide the model with a general sense of location of the pixel, e.g. if the sample is located in the valley or in the plateau. Valley Depth and VDCN provide information of how distant (vertically) is the pixel from the valley, and from the channel network. Curvatures provide general information about the terrain, and Profile Curvature can help the model to determine the stability of an area. TWI and the Natural Logarithm of Flow Accumulation show the areas that are more prone to have bodies of water, Slope and LS Factor show the steepest and more prone to erosion areas, while Aspect and Hillshade provide information about the orientation of the terrain. Statistics for the attributes are presented in Tab. 1. The maximum Elevation for the area is 997 m, and two measurements that interpret the vertical depth, Valley Depth and VDCN, have maximums between 400 and 500 m. Average slope in the area is 13.1^o, but, in the scarps area, it can get as high as 79.04^o. Planar and Profile Curvature distributions, as expected, are centered in zero.

Table 1: General statistics for each attribute, inside the study area.

Attribute	Minimum	Maximum	Average	Std. Dev.
Aspect	0.00	359.73	178.67	102.78
Elevation	15.00	997.00	543.77	312.00
Hillshade	0.00	2.29	0.96	0.44
ln of Flow Accumulation	5.05	20.55	7.98	1.90
LS Factor	0.00	122.97	3.98	3.91
Planar Curvature	-1.35	1.18	0.00	0.11
Profile Curvature	-3.17	1.70	0.00	0.13
Slope	0.00	79.04	13.10	9.82
TWI	0.59	25.23	7.24	2.93
Valley Depth	0.00	400.94	55.57	58.20
VDCN	0.00	454.17	36.66	50.66

Dimensionality reduction was not performed because prior research has shown that these eleven attributes conceive new information to the model, and removing some of them causes the model performance to drop(Lucchese et al., 2020).

3. General Sampling Procedure

Sampling of both occurrence and nonoccurrence samples was done on QGis 3.4. The points to be used as occurrence samples were acquired from inside the mapped scars polygons. To sample the maximum number of points, sampling points were located on a grid, so that each different point of the raster inside the polygons generates a sample. With this procedure, 6740 occurrence samples were generated.

We used a ratio of 50% occurrence and 50% nonoccurrence samples, in order to have a balance between them. Nonoccurrence samples were sampled inside thirteen different groups of polygons so that their location could be analyzed. Their number was held constant and equal to 6740 for all cases. The sampling procedure applied inside the polygons is a random sampling with a minimum distance of 17.7 m. This distance is the hypotenuse of a square of side 12.5 m (the raster resolution) and is set to ensure that each of the points is located over a different raster point.

4. Nonoccurrence sampling

One of the most relevant questions about nonoccurrence sampling is how distant from the landslide scar should the sample be in order for the place to be considered safe. The first set of cases comprises this aspect, consisting of six cases in which the minimum distance from the scars in nonoccurrence sample differ. This minimum distance to the scars is here called buffer.

The polygon of the buffer areas was created to have 5 km maximum distance from the scars. The minimum distance from the scars is varied, ranging from 0 m to 2.5 km. In the first group of cases, called Buffer Only or simply BO, nonoccurrence samples are acquired solely from this area. The sampling areas for each case are presented on Fig. 3. BO cases represent the logic of sampling from places close to the area where the landslides occurred because they may have had the same triggering conditions yet no landslides occurred. The rainfall (and other weather conditions) may have been more intense in the proximity of the mapped rainfall-induced landslides. Therefore, in the sampling by buffer area, restricting the samples to this radius implies in an assumption that, in this area, the rainfall was homogeneous and other terrain attributes influenced the occurrence (or not) of landslides. Our aim is that the ANN should be capable of identifying the occurrence of landslides from the input variables provided.

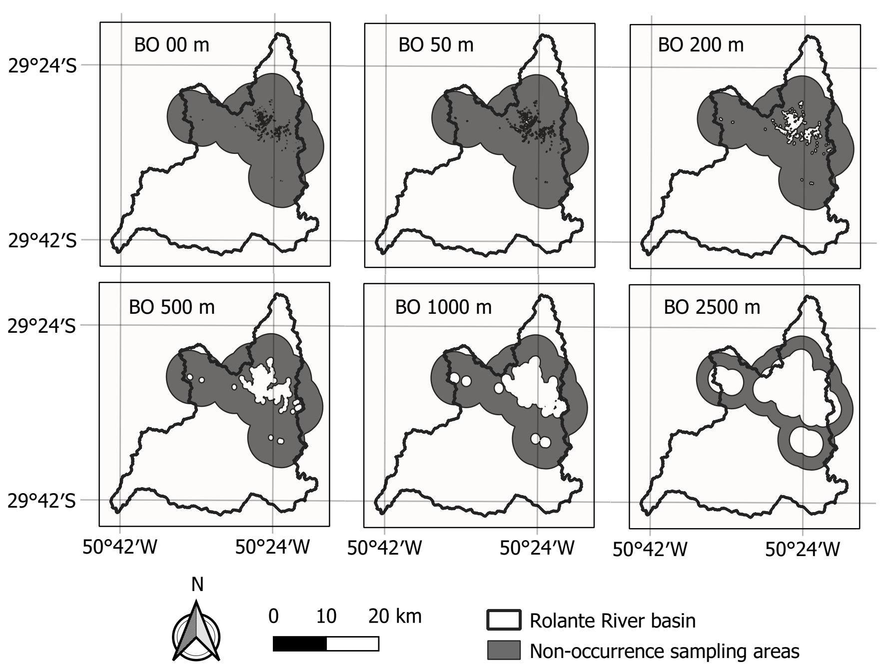

Figure 3: Buffer Only (BO) nonoccurrence sampling areas, for different minimum distances from scar area.

Another possible area for acquiring nonoccurrence samples is a rectangle in the valley area. This rectangle is shown on Fig. 4. The valley area is notably a safe region and its employment represents the nonoccurrence sampling procedure in which areas known not to be prone to landslides are taken as nonoccurrence samples. This rectangle in the lowlands has a physical meaning, because the landslides as they are a gravitational sediment movement, cannot be triggered in flat areas. Therefore, this is a way of integrating a priori knowledge in order to help the ANN model in the classification process.

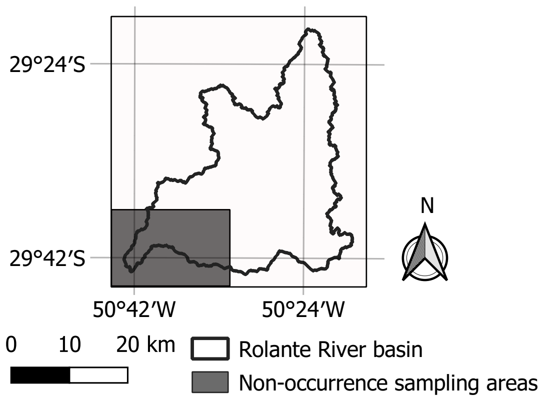

Figure 4: Rectangle Only (RO) sampling area.

The last sample group, named Buffer and Rectangle (BR), was trained considering both the buffers used for BO, and the rectangle used on RO (Fig. 5). One should notice, even if two polygons are used, the number of total nonoccurrence samples is held constant, therefore the sampling points are simply more sparse within these areas. BR group is used to analyze the effect of combining areas from BO and RO in the resulting maps and statistics. We also investigate if the maps and metrics would be close to BO or RO, or if they have their own definite characteristics.

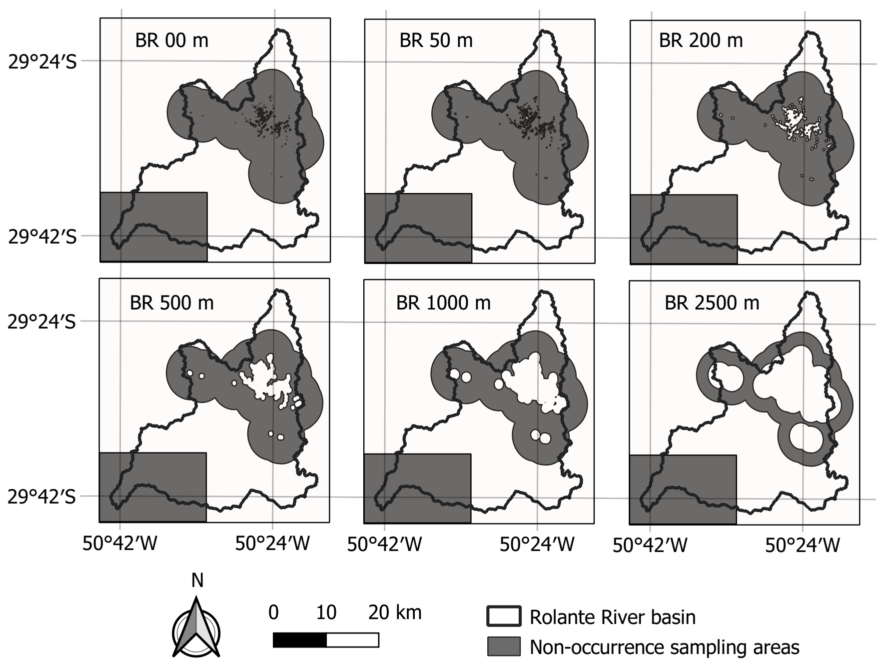

Figure 5: Buffer and Rectangle (BR) nonoccurrence sampling areas, for different minimum distances from scar area.

5. Artificial Neural Networks employed

The Artificial Neural Networks (ANNs) used are of the type Multilayer Perceptron (MLP), otherwise referred as Back-propagation Neural Network by some authors. The MLP employed consists of three layers: input, hidden, and output. The use of an ANN with one hidden layer is based on the work of Hornik et al. (1989), which stated that a neural network with a single hidden layer is able to represent any measurable relationship $r : R^n \rightarrow R^m$ if it is appropriately trained and relies on a sufficient number of neurons. Therefore, for the present case, building an ANN with more than one hidden layer is not necessary. In the hidden layer used, 30 neurons are employed. The number of neurons in the hidden layer was chosen by using an in-house developed method. The ANN should be able to perform as accurately for the validation sample as a purposefully oversized model that has been trained without overfitting. Therefore, resulting in a parsimonious model, without loss of generalization capacity when compared to models with higher complexity. This method was further described in Lucchese et al. (2020). In the input layer, the 11 attributes are used. The output, consisting of one variable, susceptibility, varies from 0 (low susceptibility) to 1 (high susceptibility).

Some commonly employed attributes were not used to train our ANNs. Land use was not employed because the region of the mapped landslide scars is uninhabited, and that could lead our model to the erroneous assumption that urban areas are not prone to landslides. Public lithology maps for this region are available only in large scales, not compatible with our area. If used, by the available classification, all scars would be located over the same lithologic unit, and so using a lithology map would not improve training in this case, as well. Instead, only attributes that can be generated from a DEM were used. This possibly makes the model and analysis more generalizable since DEMs are freely available for most of the Earth.

For the ANNs, the delta rule used is:

$$\omega_t = \omega_{t-1} + \tau\delta P_k + mo(\omega_{t-1} - \omega_{t-2}) \tag{1}$$

in which ω are the weights, τ is the learning rate, P_k is the input of layer k, and mo is the moment, set equal to 0.96 unless the error increased in the last epoch. t denotes the current epoch. δ is defined as $e_k s'_k(\eta_k)$, for which e_k is the error on layer k and $s'_k(\eta_k)$ is the activation function derivative. The activation function employed is unipolar sigmoid.

Learning rate used for training is heuristically varied and based on Vogl et al. (1988) work. The initial rate for all ANNs is $\tau = 0.00001$. If, in a given epoch, the square error rises, $\tau = 0.5\tau$, if it drops, $\tau = 1.1\tau$.

Training uses cross-validation method. For that, the original samples are divided in three: training (50%), validation (25%) and verification (25%). The 70/30 ratio between training and testing samples, commonly employed on papers of the landslide susceptibility mapping knowledge area, could not be simply applied to cross-validation training, because, in this case, three sets of samples are needed. The first set is used for training the ANN, while the second one is executed at each epoch during the training, to ensure that the ANN does not overfit to the training data. Practically, training is stopped if no improvement is made in the validation sample in 10,000 epochs, thus the weights from the last epoch when the errors on the validation sample decreased are chosen. Verification (or test) set is used mainly for metric calculation because it did not participate in training phase, therefore it is likely unbiased. The three sets must be satisfactorily representative of the whole, and the training sets used by most authors are larger than the other two. Based on that, we distributed the samples in a ratio of 50/25/25 (training/validation/verification). For each of the sets, the 50-50% ratio between occurrence and nonoccurrence samples is kept. All statistics here presented are calculated based on verification samples, that do not interfere in training.

For the sampling division, the training samples are first selected, without replacement. A random distribution is applied to acquire 50% of the occurrence and 50% of the nonoccurrence samples available, making sure the extreme values of each attribute are selected. This is done to provide a training set with as much amplitude of values as possible for the ANN training, to improve ANN generalization capability. These samples compose the training sample, which is 50% of the total samples available (realistically, it can be 50.00% - 50.16%, because of the 22 extreme values). Then, randomly, 25% of total occurrence samples, and 25% of total nonoccurrence are selected, without replacement, from the remaining samples, to compose the validation sample. The 25% of samples remaining (realistically, 24.84% - 25.00%) compose the verification sample.

For each case, five different sampling divisions between training, validation and verification are made. This is done to ensure the reliability of our analysis since the distribution of the verification samples can generate sets that have easier or harder to classify verification samples. We believe that, taking the evaluation metrics average from five divisions, the presumable variability is attenuated. The average of the verification set evaluation metrics for these five sampling divisions is calculated and presented as the final evaluation metric. The susceptibility maps presented for the cases are also the average between the maps generated by these five ANNs. Fig. 6 illustrates the five sampling divisions for one case, specifically, RO, showing the locations of training, validation and verification samples, that are visibly randomly distributed along the maps.

For the present analysis, 325 ANNs were trained. A summary of them is presented on Tab. 2. Thirteen cases for different nonoccurrence sampling

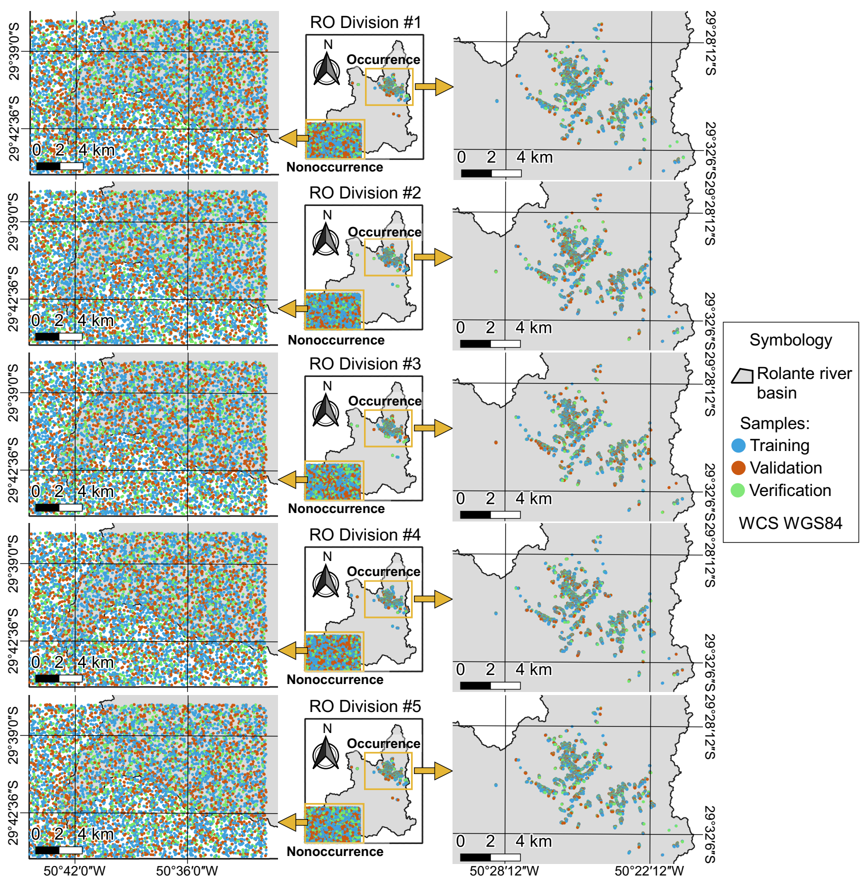

Figure 6: Location of training, validation and verification samples for each sampling division of case RO.

locations are analyzed. But, within the global samples acquired from the polygons, five different sampling divisions between training, validation and verification are made. Each of the ANNs used to calculate the average is the one chosen from five other trained ANNs with same sampling division and different initial weights. Five repetitions with different random initial weights are trained and the ANN with the best AUC on the validation sample is chosen. Because the division between training, validation and verification is constant for the five repetitions, AUC metric can be considered unbiased and the only aspect being measured is the effectiveness of the initial weights. From the five trained ANNs, one is chosen. Its evaluation metrics for verification sample are calculated and its map is generated, in order to serve as one of the five factors to calculate the general case map and metric. A flow chart for this methodology is presented on Fig. 7.

Table 2: Summary of ANNs and cases presented

	Number of cases
Sampling area	6 sizes of buffers + 6 sizes of buffers with rectangle + 1 only the rectangle = 13 types of areas
Sampling division	5 random sampling divisions for each
Initial weights	5 repetitions of training with different initial weights
Total number of ANNs trained	13*5*5 = 325

Our in-house ANN code was fully developed on Matlab platform. Earlier versions of this algorithm were employed in the papers of Fantin-Cruz et al. (2011), Dornelles et al. (2013), Oliveira et al. (2015), Moreira de Melo and Pedrollo (2015) and Sari et al. (2017).

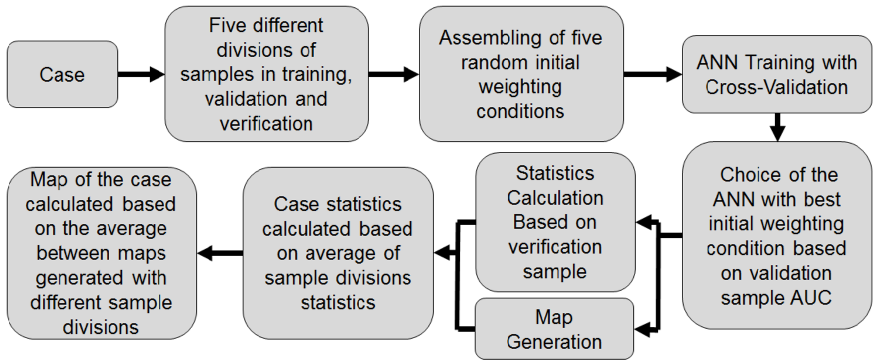

Figure 7: Flow chart of the ANNs presented on the present paper.

6. Evaluation metrics and analysis

In this paper, two main evaluation metrics are presented and analyzed, accuracy and AUC (DeLong et al., 1988).

Accuracy is the rate of right answers provided by the model, or

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

in which TP is True Positive (model predicted TRUE, it was a hit), TN is True Negative (model predicted FALSE, it was a hit), FP is False Positive (model predicted TRUE, missed it), and FN is False Negative (model predicted FALSE, missed it). The accuracy measures the ability of the model to predict the right (boolean) answer, based on the samples provided. Our ANN model output is a continuous value between 0 (low susceptibility) and 1 (high susceptibility). To calculate variables TP, TN, FP, and FN, a threshold over which the output is considered susceptible must be established. Instead of choosing an arbitrary threshold, its value is varied from 0 to 1, and accuracies are calculated for each threshold within the limits. The threshold that provides the highest accuracy is chosen, usually gravitating around 0.5. Accuracy is one of the main evaluation metrics used to evaluate models, and, generally, it is considered that, the closer to 1, the better.

AUC is the area under the Receiver Operator Characteristic (ROC) curve, which is True Positive Rate (TPR) by False Positive Rate (FPR), for a range of thresholds. TPR is $TP/(TP+FN)$ and FPR is $FP/(FP+TN)$. Generally, models present an AUC between 0.5 and 1, with models closer to 1 usually being considered more reliable.

In this paper, the evaluation metrics *acc* and AUC presented in the following section always refer to the ones calculated based on the verification sample. The same model, applied to different samples, generates different *acc* and AUC values.

7. Results

In this section, results for the 13 cases are presented. Evaluation metrics AUC and *acc*, calculated based on verification samples, are shown. The resulting landslide susceptibility map for each case is also presented.

In Tab. 3, the average of the verification sample AUC and accuracy *acc* for different sampling divisions is presented for each case. AUC varies from 0.94 to 1, while accuracies range from 89% to 100%. Evaluation metrics based on the Buffer Only (BO) sampling are lower than those from Buffer and Rectangle (BR). Rectangle Only (RO) presented the best evaluation metrics of all, for which AUC is 1 and accuracy is 100%, indicating all the models trained achieved 100% of right classifications. These evaluation metrics were calculated based on the verification samples presented on Fig. 6, and averaged for all five divisions. For the RO case, in all divisions, FN and FP are zero, and TP and TN are at their maximum values, for a given threshold. It also means that as TPR may be written as $TP/(TP+0)$, if $TP \neq 0$, this results in $TPR = 1$. Analogously, FPR is $0/(0+TN) = 0$, therefore, for $TN \neq 0$, $FPR = 0$. This occurred because all the occurrence samples were in the scarped areas and all the nonoccurrence samples are in the rectangle on the valley (in the training, validation and verification sets), resulting on the fact that the model reproduced this pattern. Minimum distances to scars influences the resulting evaluation metrics values. Between cases of the same group, the larger this distance, the higher the AUC and the *acc*.

Table 3: Summary of cases and their verification sample AUC and acc.

		AUC	acc
Buffer Only (BO)	BO 00m	0.9409	89.45%
	BO 50m	0.9494	90.26%
	BO 200m	0.9524	90.67%
	BO 500m	0.9569	91.18%
	BO 1000m	0.9586	91.24%
	BO 2500m	0.9662	92.28%
Buffer and Rectangle (BR)	BR 00m	0.9616	92.33%
	BR 50m	0.9647	92.28%
	BR 200m	0.9682	93.05%
	BR 500m	0.9686	92.96%
	BR 1000m	0.9680	93.19%
	BR 2500m	0.9806	95.09%
Rectangle Only (RO)		1.0000	100.00%

Part of our analysis consists of the observation of the susceptibility maps generated. Thus, it is necessary to plot maps for each case. These maps are simple averages of the maps generated from the five different sample divisions (in training, validation and verification, see Tab. 2). In this paper, for the maps shown, the range of susceptibility is continuous, as the output is provided by the ANN. No classification or alteration was performed prior to map plotting.

In Fig. 8, resulting susceptibility maps are shown for the six cases comprised on the Buffer Only non-ocurrence sampling. All the six maps generated using this nonoccurrence sampling procedure are satisfactory and tend to present scarped areas as susceptible. BR 200m, BR 500m and BR 2500m show non-null susceptibility values in places in the top of the mountain, generally considered safe.

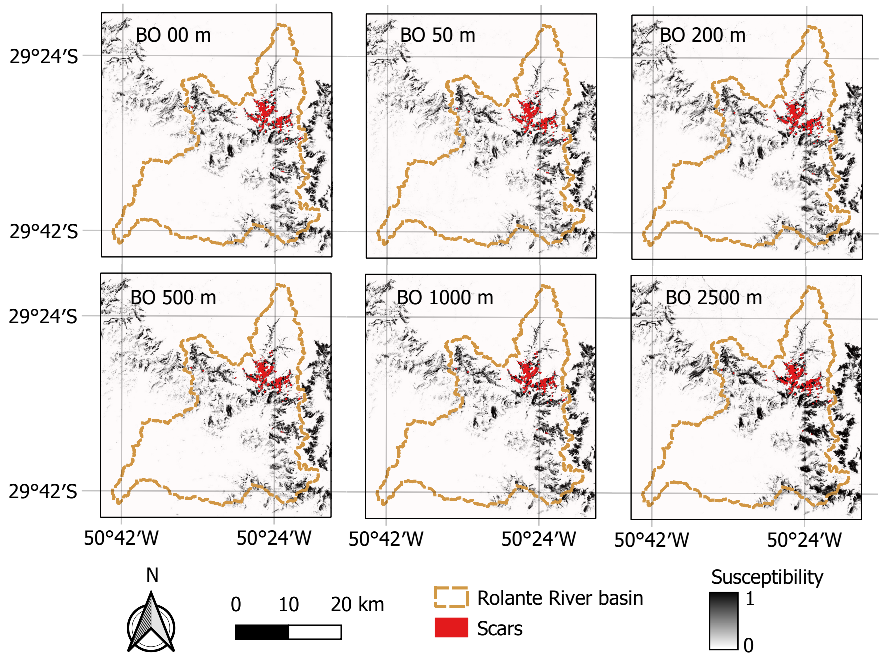

Figure 8: Buffer Only (BO) resulting susceptibility maps.

Hazard susceptibility maps were also plotted for Buffer and Rectangle cases (Fig. 9). They are very alike the maps generated for cases of group BO. Some of these maps also present diversions on the mountain top area.

A susceptibility map was generated based on the models of case Rectangle Only, that considers only an area in the valley as non-susceptible (Fig. 10). According to this map, the plateau area would be considered susceptible, which is known to be a misclassification. Comparing Fig. 10 to Fig. 6, it is possible to see that the nonoccurrence samples (for training, validation and verification) are contained in the area classified as not susceptible, and the occurrence samples (for training, validation and verification) are contained in the area classified as susceptible. Therefore, they were all correctly classified, corroborating the

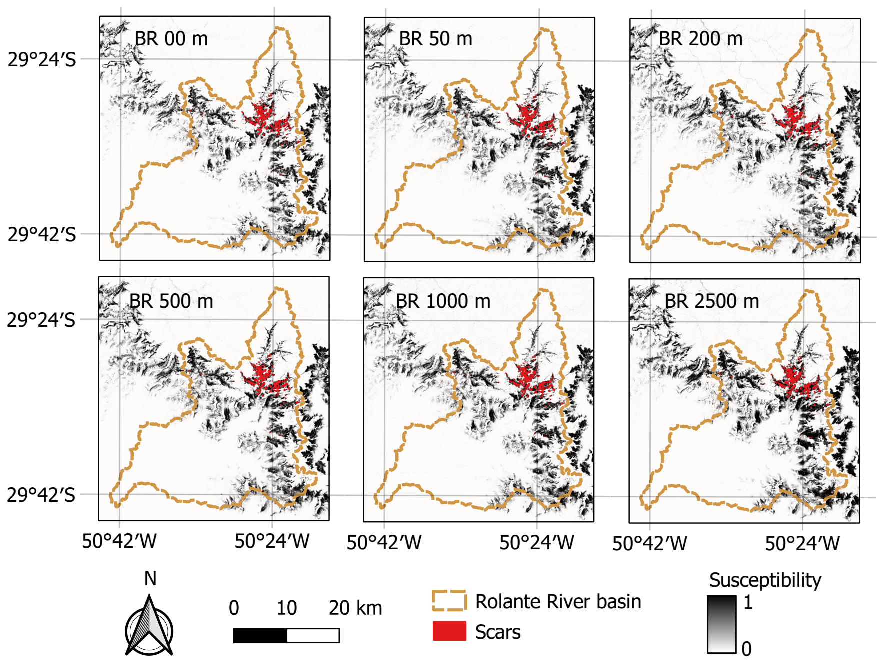

Figure 9: Buffer and Rectangle (BR) resulting susceptibility maps.

evaluation metrics AUC=1 and acc=100% presented in Tab. 3. It is the least constrained case, as well, and would not be considered suited for zonation because a vast area is considered susceptible.

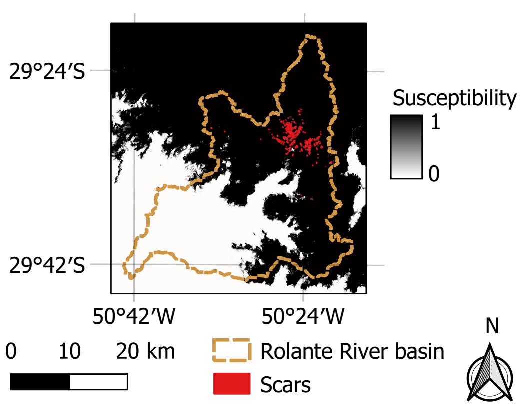

Figure 10: Rectangle Only (RO) resulting susceptibility map.

8. Discussions

If the criterion to select not susceptible areas was similar to the one used by Gomez and Kavzoglu (2005); Choi et al. (2010), the selection of nonoccurrence samples would be based on selecting places known not to be prone to landslides, such as the valley area. For this reason, this sampling would be somewhat similar to using a rectangle in the valley area, such as done on case Rectangle Only (RO). As observed, this sampling provided the best evaluation metrics, since every sample was correctly classified. If we were to choose between nonoccurrence sampling methods based only on the evaluation metrics analyzed, we would point RO case as the best option. However, it should be noted that the evaluation metrics are not suited for direct comparisons in this case, because the verification samples composition is different. Also, when the map is observed, and, specially, when it is confronted with maps generated with samples based on other procedures, the difference is clear. The map generated by RO is not constrained, it classifies a vast area as susceptible, including the plateau area which is known for not being susceptible to landslides. That may be considered, depending on the point of view, as incompatible with reality. It would as well not be suited for hazard zonation for this motive. The reason it has high evaluation metrics has no connection with being more precise - in fact, it is the opposite - the samples provided as a dataset for training are too easy to classify. They are so distant from each other, and so different in many aspects that even a 'loose', 'careless' classification would classify every single one of them correctly. In other words, it is not hard for the model to achieve 100% accuracy because the samples themselves are biased. This may also be the reason why, in Gomez and Kavzoglu (2005), high accuracy was achieved, even if the map shown seems to reflect an interpretation of the terrain Elevation attribute.

Zhu et al. (2019) also observed this effect. Models based on nonoccurrence samples that are very different from the occurrence samples provide higher accuracies, although the maps generated based on them roughly overestimate occurrence areas. This resonates to what is observed by the present authors. Hong et al. (2019) results showed that sampling methods that acquire nonoccurrence samples in a similar setting but farther from occurrence samples result in maps with larger susceptible areas, also in agreement with the present analysis. This is not necessarily wrong or provides worse maps, although excessive intervention of human expert knowledge-based processes on dataset forming hinders the possibilities for the ANNs to find their own ways.

The evaluation metrics were calculated to test the intrinsic adjust of each model, in which the RO case resulted in a perfect adjust to the samples that were provided to it. However, notably, the evaluation metrics hereby cited are not calculated based on the same verification sample, as they would be in an ideal setting. That said, the same effect discussed for RO can be observed with the increase on the minimum distance (internal buffer) on which the sample is acquired. Even if all BO cases are generating similar maps, the farther from the scars these samples are, the easier they are to classify. On BR samples, the evaluation metrics of the models were in between the values of BO and RO, but the maps looked much more like BO maps. It is possible to think that the easily classifiable samples, provided by the rectangle on the lowlands, made it easier for the model to classify the verification set, which is acquired from the same sample pool as the training and validation sets for each case, and this pool had now easy-to-classify samples. We should remark that easier-to-classify does not equate directly to being better because it does not force the model to train more in order to discriminate between samples that are alike, but produce different outcomes, and purposeful comparisons can only be done if the metrics are calculated based on similar samples.

As observed by Zhu et al. (2018), the existence itself of the nonoccurrence samples in the calibration/training of a method has a remarkable effect on its resulting maps. They also discussed that methods that use nonoccurrence samples tend to constrain the over-prediction of high susceptibility. With the analyses here presented, we show that not only their presence is important, but their location and their acquisiton process are as well relevant. Many previous works had not presented the location neither commented on the method used to extract nonoccurrence samples, and yet, as we observe, this information is relevant to interpret the resulting evaluation metrics correctly.

9. Conclusion

In the present paper, we have provided answers to pressing questions regarding nonoccurrence samples used to train ANN models for landslide susceptibility assessment. Using 13 cases with different locations for the nonoccurrence samples, 325 ANNs were trained to provide a reliable outlook on this subject.

For the conducted analyses, AUC and accuracy were chosen as example evaluation metrics because they are two of the most commonly used. Although more research in this area is needed for all possible methods for nonoccurrence sampling to be contemplated, we showed that the locations of these samples are very relevant, with visible effects on the generated map. By intervening and choosing nonoccurrence samples that are distant and have very different attributes from the occurrence ones, we have shown that, using this configuration, ANN, as it is a data-driven model, is not capable of acquiring the necessary knowledge in order to correctly discriminate between susceptible and non-susceptible areas. This is possible to observe in the maps generated, which are not constrained. A probable explanation to why occurrence and nonoccurrence samples used to train the models tend to be different in these settings is that the classification human knowledge allows us to do (such as choosing the lowlands as the nonoccurrence area) is likely not thorough enough. This is one of the reasons why researchers usually do not rely solely on human perception for susceptibility mapping, using models instead. Even if, in this case, AUC and *acc* metrics are maximized, because all classifications are right, this happens because the sample has a biased configuration that makes it too easy for the model to provide the right output.

Expert knowledge is very important to landslide-related studies. However, in many cases, this expert knowledge intervention should, instead of being applied on the formulation of the sample, (e.g. including a rectangle on the lowlands), be applied to the delimitation of the application areas of the final ANN model. Therefore, a possible course of action would be admitting a priori that the flat regions are not susceptible and, based on that, not generating the maps with the model in these regions, if a problem in which the model has not captured the knowledge of how to classify the lowlands is observed. Another way to put it, one can say that what is meant with the inclusion of easily classifiable areas is achievable anyway by defining the application areas possible to be mapped by the model.

In many previously published papers, the locations and methods of acquisition of nonoccurrence samples are not presented or commented. However, we have shown that these are important aspects to be considered when analyzing the evaluation metrics of an ANN model for landslide susceptibility. Overall, based on the empirical evidences presented, we can state that using easily classifiable samples, the model may present high accuracy and AUC, without necessarily this equating to the generation of generalizable map.

10. Acknowledgments

This research was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) Edital 01/2017 - ARD, process 17/2551-0000894-4.

References

Aleotti, P., Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. Bulletin of Engineering Geology and the environment 58, 21–44.

Anbalagan, R., 1992. Landslide hazard evaluation and zonation mapping in mountainous terrain. Engineering geology 32, 269–277.

Arnone, E., Francipane, A., Scarbaci, A., Puglisi, C., Noto, L.V., 2016. Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. Environmental modelling & software 84, 467–481.

ASF DAAC, 2015. Alos palsar radiometric terrain corrected high res. https://www.asf.alaska.edu. doi:10.5067/Z97HFCNKR6VA. includes Material (c) JAXA/METI 2006-2011. Accessed through ASF DAAC 29 December 2018.

Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65, 15–31.

Braun, A., Urquia, E.L.G., Lopez, R.M., Yamagishi, H., 2019. Landslide susceptibility mapping in Tegucigalpa, Honduras, using data mining methods, in: IAEG/AEG Annual Meeting Proceedings, San Francisco, California, 2018-Volume 1, Springer. pp. 207–215.

Bui, D.T., Pradhan, B., Lofman, O., Revhaug, I., 2012. Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models. Mathematical problems in Engineering 2012.

Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13, 361–378.

Can, A., Dagdelenler, G., Ercanoglu, M., Sonmez, H., 2019. Landslide susceptibility mapping at Ovacık-Karabük (Turkey) using different artificial neural network models: comparison of training algorithms. Bulletin of Engineering Geology and the Environment 78, 89–102.

Catani, F., Lagomarsino, D., Segoni, S., Tofani, V., 2013. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Natural Hazards and Earth System Sciences 13, 2815–2831.

Chen, W., Pourghasemi, H.R., Zhao, Z., 2017a. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. Geocarto international 32, 367–385.

Chen, W., Shirzadi, A., Shahabi, H., Ahmad, B.B., Zhang, S., Hong, H., Zhang, N., 2017b. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. Geomatics, Natural Hazards and Risk 8, 1955–1977.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z., Ma, J., 2017c. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena 151, 147–160.

Choi, J., Oh, H.J., Won, J.S., Lee, S., 2010. Validation of an artificial neural network model for landslide susceptibility mapping. Environmental Earth Sciences 60, 473–483.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845.

Dornelles, F., Goldenfum, J.A., Pedrollo, O.C., 2013. Artificial neural network methods applied to forecasting river levels. Revista Brasileira de Recursos Hídricos 18, 45–54.

Dou, J., Yamagishi, H., Pourghasemi, H.R., Yunus, A.P., Song, X., Xu, Y., Zhu, Z., 2015. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. Natural Hazards 78, 1749–1776.

Dou, J., Yamagishi, H., Zhu, Z., Yunus, A.P., Chen, C.W., 2018. TXT-tool 1.081-6.1 a comparative study of the Binary Logistic Regression (BLR) and Artificial Neural Network (ANN) models for GIS-based spatial predicting landslides at a regional scale, in: Landslide Dynamics: ISDR-ICL Landslide Interactive Teaching Tools. Springer, pp. 139–151.

Dou, J., Yunus, A.P., Bui, D.T., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.W., Khosravi, K., Yang, Y., Pham, B.T., 2019. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Science of the Total Environment 662, 332–346.

Ercanoglu, M., Gokceoglu, C., 2002. Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach. Environmental geology 41, 720–730.

Ermini, L., Catani, F., Casagli, N., 2005. Artificial neural networks applied to landslide susceptibility assessment. Geomorphology 66, 327–343.

Fall, M., Azzam, R., Noubactep, C., 2006. A multi-method approach to study the stability of natural slopes and landslide susceptibility mapping. Engineering geology 82, 241–263.

Fantin-Cruz, I., Pedrollo, O., Castro, N.M., Girard, P., Zeilhofer, P., Hamilton, S.K., 2011. Historical reconstruction of floodplain inundation in the Pantanal (Brazil) using neural networks. Journal of Hydrology 399, 376–384.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z., et al., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. Engineering Geology 102, 99–111.

García-Rodríguez, M., Malpica, J., 2010. Assessment of earthquake-triggered landslide susceptibility in El Salvador based on an Artificial Neural Network model. Natural Hazards and Earth System Sciences 10, 1307–1315.

Gökceoglu, C., Aksoy, H., 1996. Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. Engineering Geology 44, 147–161.

Gomez, H., Kavzoglu, T., 2005. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. Engineering Geology 78, 11–27.

Gong, Q.h., Zhang, J.x., Wang, J., 2018. Application of GIS-based back propagation artificial neural networks and logistic regression for shallow landslide susceptibility mapping in South China- take Meijiang River Basin as an example. The Open Civil Engineering Journal 12.

Gutiérrez-Martín, A., 2020. A gis-physically-based emergency methodology for predicting rainfall-induced shallow landslide zonation. Geomorphology , 107121.

Hong, H., Miao, Y., Liu, J., Zhu, A.X., 2019. Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. Catena 176, 45–64.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural networks 2, 359–366.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P., Stocker, E.F., 2007. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. Journal of hydrometeorology 8, 38–55.

Hussin, H.Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., Bălteanu, D., 2016. Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. Geomorphology 253, 508–523.

Kanungo, D., Arora, M., Sarkar, S., Gupta, R., 2006. A comparative study of conventional, ann black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Engineering Geology 85, 347–366.

Kawabata, D., Bandibas, J., 2009. Landslide susceptibility mapping using geological data, a dem from ASTER images and an Artificial Neural Network (ANN). Geomorphology 113, 97–109.

Kumar, D., Rawat, A., et al., 2018. Study and prediction of landslide in

Uttarkashi, Uttarakhand, India using GIS and ANN. American Journal of Neural Networks and Applications 3, 63.

Lee, S., 2005. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. International Journal of Remote Sensing 26, 1477–1491.

Lee, S., Min, K., 2001. Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental geology 40, 1095–1113.

Lee, S., Ryu, J.H., Won, J.S., Park, H.J., 2004. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Engineering Geology 71, 289–302.

Lucchese, L.V., Oliveira, G.G.d., Pedrollo, O.C., 2020. Attribute selection using correlations and principal components for artificial neural networks employment for landslide susceptibility assessment. Environmental Monitoring and Assessment 192, 129.

Moreira de Melo, T., Pedrollo, O.C., 2015. Artificial neural networks for estimating soil water retention curve using fitted and measured data. Applied and Environmental Soil Science 2015.

Merghadi, A., Abderrahmane, B., Tien Bui, D., 2018. Landslide susceptibility assessment at Mila Basin (Algeria): a comparative assessment of prediction capability of advanced machine learning methods. ISPRS International Journal of Geo-Information 7, 268.

Oliveira, G.G., Pedrollo, O.C., Castro, N.M., 2015. Simplifying artificial neural network models of river basin behaviour by an automated procedure for input variable selection. Engineering Applications of Artificial Intelligence 40, 47–61.

Ortiz, J.A.V., Martínez-Graña, A.M., 2018. A neural network model applied to landslide susceptibility analysis (Capitanejo, Colombia). Geomatics, Natural Hazards and Risk 9, 1106–1128.

Pham, B.T., Bui, D.T., Prakash, I., Dholakia, M., 2017. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. Catena 149, 52–63.

Polykretis, C., Chalkias, C., 2018. Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models. Natural hazards 93, 249–274.

Pourghasemi, H.R., Kerle, N., 2016. Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. Environmental earth sciences 75, 185.

Pradhan, B., Lee, S., 2010. Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. Environmental Earth Sciences 60, 1037–1054.

Regmi, A.D., Devkota, K.C., Yoshida, K., Pradhan, B., Pourghasemi, H.R., Kumamoto, T., Akgun, A., 2014. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. Arabian Journal of Geosciences 7, 725–742.

Sari, V., dos Reis Castro, N.M., Pedrollo, O.C., 2017. Estimate of suspended sediment concentration from monitored data of turbidity and water level using artificial neural networks. Water Resources Management 31, 4909–4923.

Secretaria Estadual do Meio Ambiente, Grupo de Pesquisa em Desastres

Naturais, 2017. Diagnóstico preliminar. Technical Report. Departamento de Recursos Hídricos da Secretaria Estadual do Meio Ambiente (DRH-SEMA). `http://www.rs.gov.br/upload/20170125183225diagnostico_preliminar_gt_rolante_revfinal.pdf`.

Shirzadi, A., Solaimani, K., Roshan, M.H., Kavian, A., Chapi, K., Shahabi, H., Keesstra, S., Ahmad, B.B., Bui, D.T., 2019. Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. Catena 178, 172–188.

Sorriso Valvo, M., 2002. Landslides; from inventory to risk, in: Landslides Proceedings of the International European Conference on Landslides. Balkema, Rotterdam, pp. 79–93.

Süzen, M.L., Doyuran, V., 2004. Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. Engineering geology 71, 303–321.

Varnes, D.J., 1984. Landslide hazard zonation: a review of principles and practice. Natural Hazards 3.

Vogl, T.P., Mangis, J., Rigler, A., Zink, W., Alkon, D., 1988. Accelerating the convergence of the back-propagation method. Biological cybernetics 59, 257–263.

Wang, L.J., Sawada, K., Moriguchi, S., 2013. Landslide susceptibility analysis with logistic regression model based on fcm sampling strategy. Computers & Geosciences 57, 81–92.

Xiao, C., Tian, Y., Shi, W., Guo, Q., Wu, L., 2010. A new method of pseudo absence data generation in landslide susceptibility mapping with a case study of shenzhen. Science China Technological Sciences 53, 75–84.

Xiong, J., Sun, M., Zhang, H., Cheng, W., Yang, Y., Sun, M., Cao, Y., Wang, J., 2019. Application of the Levenburg–Marquardt back propagation neural network approach for landslide risk assessments. Natural Hazards and Earth System Sciences 19, 629–653.

Yao, X., Tham, L., Dai, F., 2008. Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. Geomorphology 101, 572–582.

Yesilnacar, E., Topal, T., 2005. Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Engineering Geology 79, 251–266.

Yilmaz, I., 2010. The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks. Environmental Earth Sciences 60, 505–519.

Zare, M., Pourghasemi, H.R., Vafakhah, M., Pradhan, B., 2013. Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arabian Journal of Geosciences 6, 2873–2888.

Zhu, A.X., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T., Hong, H., 2019. A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods. Catena 183, 104188.

Zhu, A.X., Miao, Y., Yang, L., Bai, S., Liu, J., Hong, H., 2018. Comparison of the presence-only method and presence-absence method in landslide susceptibility mapping. Catena 171, 222–233.