Journal of Urban and Environmental Engineering, v.18, n.2, p.78-84

ISSN 1982-3932
doi: 10.4090/juee.2024.v18n2.078084

Journal of Urban and Environmental Engineering

www.journal-uee.org

MACHINE LEARNING IN THE PREDICTION OF PARAMETERS OF THE HEAVY RAINS EQUATION. CASE STUDY: NORTHWEST FLUMINENSE

Alex Tavares Silva[1*], Jader Lugon Junior[1], Pedro Paulo Gomes Watts Rodrigues[2], Vicente de Paulo Santos de Oliveria[1] and Wagner Rambaldi Telles[1]

[1] *Instituto Federal Fluminense, Brazil*

[2] *Universidade do Estado do Rio de Janeiro, Brazil*

Received 22 October 2023; received in revised form 16 May 2024; accepted 18 May 2024

Abstract: The heavy rain forecasts are important in disaster prevention as well as agricultural planning. In this paper parameters of the heavy rainfall equation were estimated using a machine learning model applying the random forests technique and with a case study in the region of Noroeste Fluminense, Brazil. The same parameters were also estimated, using the Levenberg-Marquardt optimization method, using data from satellite images. Both the Machine Learning parameters predicted by the Levenberg-Marquard model are compared with the parameters found in the literature.

Keywords: Heavy rains, Computational intelligence, Random Forest, Levenberg-Marquardt

[*] Correspondence to: Alex Tavares Silva. E-mail: altasilva@gmail.com

INTRODUCTION

Climate change has a direct impact on the increase in the intensity and recurrence of rainfall, and this is influenced by human action in nature. Botzen *et al.* (2022) understand that the public, and especially the middle class, is keen on acting against the detrimental scenarios of catastrophic climate change becoming real. According to Silva *et al.* (2021) when rain hits a surface, depending on its intensity, it can cause serious disruption to the region.

One of the effects of climate change can be seen in agriculture, which is fundamental for society because it is through it that we obtain the necessary resources for our sustenance. It is extremely important to carry out studies to try to predict both periods of drought and periods of rain. In both periods agriculture can be harmed. According to Nketia *et al.* (2022), in the last decades, machine learning (ML) approaches offer powerful tools in overcoming bottlenecks in conventional mapping approaches, becoming central in predictive soil mapping studies.

Machine learning can be applied in several types of studies, including landslides, which in China are highly dangerous and have a high frequency of occurrence (Chang and Wan, 2015, Borrelli *et al*., 2018). However, worldwide, they have not yet developed a unified standard for expressing landslide boundaries (Hong *et al*., 2018, Chang *et al*., 2021, Huang *et al*., 2020).

The parameters of the heavy rainfall equations are defined by monitoring data by meteorological stations at the place of interest. Satellite images are very important to estimate rainfall parameters, due to the precision in the data and, combined with optimization methods such as Levenberg-Marquardt (LM), can provide more accurate parameters. The Levenberg- Marquardt method is a method that fits into what are called inverse problems, which according to Fu *et al.* (2022), arise from applications in natural sciences, engineering and medical sciences whenever unknown causes are sought based on observations of their effects.

Predicting parameters of the rainfall equation in regions with scarcity of data can directly help in agriculture, since it is possible to predict periods with higher and lower rainfall. In flooding situations, several applications are available to assist in flood prevention (Telles *et al*., 2016; Tavares *et al*., 2018; Junior *et al*., 2019). Hudnurkar & Rayavarpu (2022) state that many researchers have been used by machine learning algorithms such as k-nearest neighbor (k-NN), support vector machine (SVM), artificial neural network (ANN), decision tree (DT) and random forest (RF) for the purpose of rainfall classification.

Particularly, random forest is a technique used in modeling predictions and behavior analysis and is built on decision trees. Lee *et al*. (2022) compared several models of machine learning to estimate of rainfall erosivity factor in Italy and Switzerland and concluded the results RF model had the highest performance.

Thus, the objective of this work is to develop a methodology capable of estimating the parameters of the equation of intense rains, through remote sensing (satellite), applying in the city of Varre-Sai, Rio de Janeiro, Brazil, using a method of optimization and computational intelligence to make predictions in regions with little or no data.

MATERIAL AND METHODS

Study region and data

Oliveira (2019) conducted a survey of all intensity-duration-frequency relationships of extreme rainfall published in Brazil, using data from a Brazilian regulatory agency known as Agência Nacional de Águas e Saneamento Básico (ANA), estimating parameters from 5209 rainfall stations containing more than 15 years of observations. daily, without fail. The calculated parameters are dimensionless parameters of the general rainfall equation (Tucci *et al*., 2015, Oliveira, 2019), according to Eq. 1 (general equation of heavy rains)

$$i = \frac{KTR^{a}}{\left(t+b\right)^{c}} \qquad (1)$$

where: i is standard deviation (mm); K, a, b, c are dimensionless parameters that depend on the region (mm); TR is recurrence time (mm); and t is duration (min).

The data used for the machine learning model to predict the parameters were taken from Oliveira (2019). Python programming language and scikit-learn library was used to implement the machine learning model. Scikit-learn is a free software machine learning library for the Python programming language. For create the map was used the software QGIS, version 3.10.4.

The data used to train and validate the model is shown in **Table 1**. Parameters from 6 municipalities around the municipality of Varre-Sai were used. Castelo and Dores do Rio Preto belong to the State of Espírito Santo (ES). Cardoso Moreira, Italva, Itaperuna and Porciúncula belong to the State of Rio de Janeiro (RJ).

In **Fig. 1** the map of the region of study is shown, as well as its latitudinal and longitudinal coordinates in decimal degrees.

Table 1. Data from trained/validated stations

Local	State	Latitude	Longitude	K	a	b	c
Castelo	ES	–20°36′20″	–41°11′59″	733.2717	0.1398	9.7895	0.7243
Dores do Rio Preto	ES	–20°41′09″	–41°50′46″	816.0914	0.1396	9.7859	0.7242
Cardoso Moreira	RJ	–21°29′31″	–41°36′49″′	694.0401	0.1545	9.7819	0.7242
Italva	RJ	–21°25′00″	–41°42′00″	747.7179	0.1414	9.7841	0.7242
Itaperuna	RJ	–21°12′23″	–41°53′28″	808.4799	0.1531	9.7838	0.7242
Porciúncula	RJ	–20°58′09″	–42°03′06″	792.1848	0.1433	9.7818	0.7241

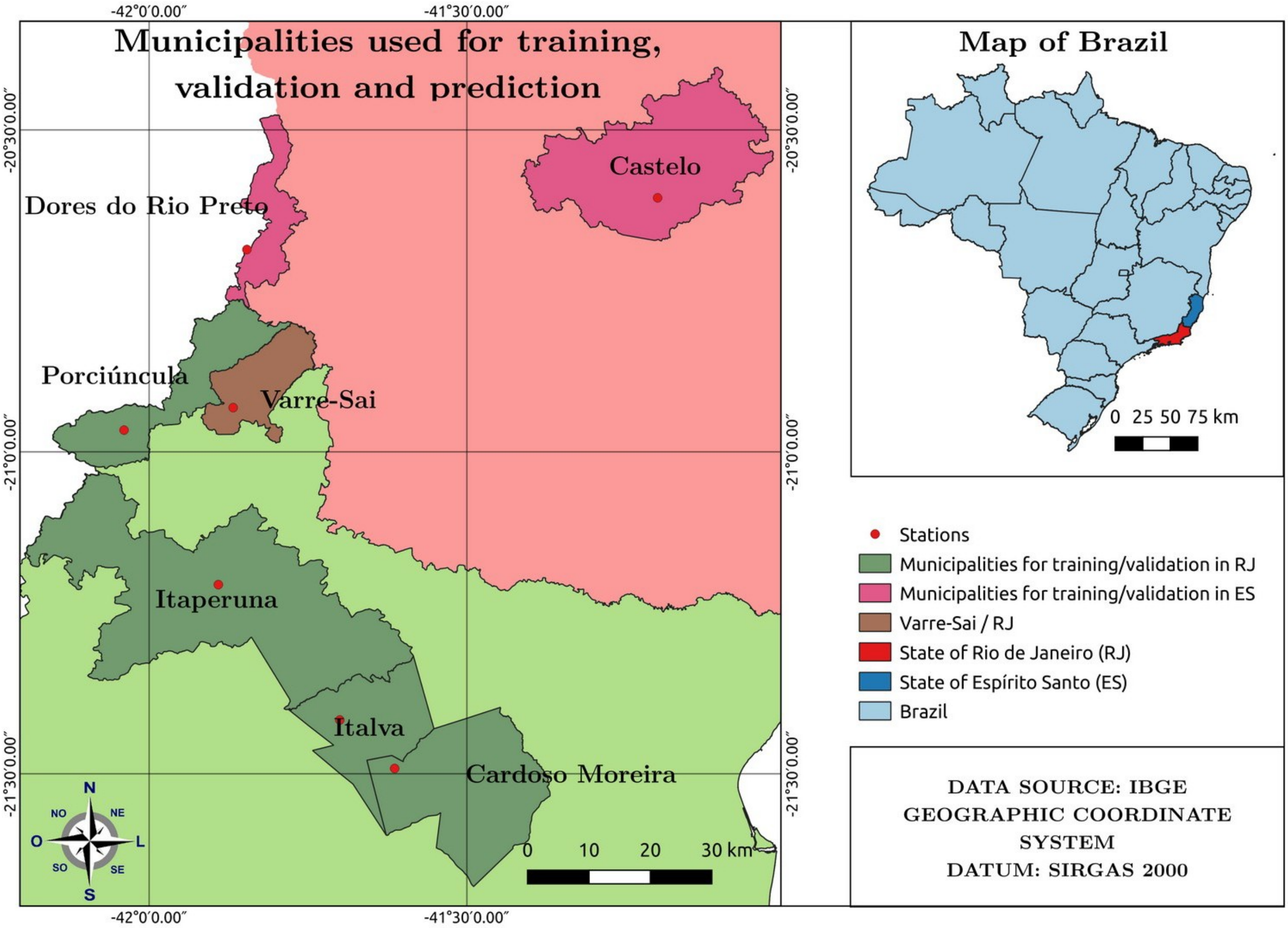

Fig. 1 Map of the study region with respective municipalities.

Machine learning model

To unite the efficiency of standard statistical techniques with the detail and complexity of physical approaches, machine learning regression algorithms (non-parametric algorithms) are able to efficiently process an expanding volume of Earth observation data (Verrelst *et al*., 2015, Verrelst *et al*., 2012). Timely and cost-effective solutions are delivered through tree-based machine learning models based on environmental resources (Li *et al*., 2022).

The algorithm used is Random Forest, a commonly used machine learning algorithm that combines the output of multiple decision trees to achieve a single result. As stated by Hijiao *et al*. (2022) the basic idea of the RF model is to extract repeatedly and randomly K samples from the training dataset with replacement according to the bootstrap resampling method.

The number of trees in the forest was equal to 15 and used the mean squared error to variance reduction as feature selection criterion. The number of samples required to split an internal node was equal 2. The number of samples to be at a leaf node was equal 1. For training, 80% of the data were used while 20% were used for validation, thus 4 municipalities for training and 2 for validation. Data for training and validation were selected randomly. The machine learning algorithm using the random forest method was performed on a total of 100 runs.

Levenberg-Marquardt algorithm

The Levenberg-Marquardt algorithm is a deterministic optimization method used to estimate parameters, that in the case of this study, are the parameters for rainfall equation. Parameter extraction is an important part of model development, and the goal of parameter extraction and optimization is to determine the values of device model parameters that minimize the differences between a set of measured characteristics and results obtained by evaluations of the device model. This minimization process is often called fitting of model characteristics to the measurement data (Duc-Hung *et al*., 2012).

The Levenberg-Marquardt algorithm was executed in a total of 100 runs and the parameters were averaged. The maximum number of iterations of the method is equal to 10 and the difference between two estimates is 1E-1.

RESULTS AND DISCUSSIONS

In the Machine Learning, for each run, the algorithm retrained and validated the data, calculating the mean absolute error (MAE). At the end of the 100 corrections, the average for each parameter was calculated, in addition to the MAE average. **Table 2** shows the lowest and highest value, the mean, the standard deviation and MAE of validation for the parameters K, a, b and c, of the 100 runs.

Due to the low standard deviation, the training/validation part of the algorithm occurred satisfactorily. Parameter c is a parameter that changes little, so it is possible to notice that the validation generates a standard deviation of zero in relation to the mean. In **Fig. 2** shows the results of the parameter prediction, as well as the results from literature and the results of the optimization algorithm using satellite data. The parameters estimated by Oliveira based on ANA data are quite different from those obtained in this work, based on satellite images. This difference is due to an error at the time of recording the rainfall. Satellite images are much more accurate.

Still in **Fig. 2** the mean absolute error in the 100 runs, in the Machine Learning algorithm, where a satisfactory error is noticed again highlighting the parameter c, which had a difference only in the fourth decimal place. With the MAE metric, the absolute value of each error is taken, converting each error into a positive number. Then, the average of these absolute errors is taken, which will be the quality measure of the model (**Table 3)**. In **Fig. 3**, it is perceived that the forecast for all parameters was satisfactory, being very close to the values of the station of the municipality of Varre-Sai estimated by Oliveira (2019). Parameter c is clearly the one that varies the least in the results because all stations have practically the same value for that parameter. In **Fig. 4** shows the intensity profiles of the 24- hour rainfall for the recurrence times of 2, 5, 10, 20, 50 and 100 years.

Table 2. Results of the validation

	Validation			
	K	a	b	c
Mean	765.7490	0.1409	9.7867	0.7242
Standard Deviation	9.2267	0.0008	0.0006	0.0000
Lowest	745.5346	0.1398	9.7854	0.7242
Highest	789.4479	0.1436	9.7881	0.7243
MAE Mean	53.5314	0.0070	0.0045	0.0001

CONCLUSIONS

1. The method is suitable for predicting the parameters of the heavy rainfall equation in places without a meteorological station.

2. The results improved when the number of trees in the forest was adjusted to 15. Above or below this value or below, the results tend to deviate from the expected value.

3. After 100 runs the validation and prediction means were satisfactory. In the validation due to the fact that the MAE has a satisfactory value and in the prediction due to the comparison with the results found in the literature.

Table 3. Results of the prediction of ML compared to LM

Varre-Sai / RJ	Prediction							
	Machine Learning				Levenberg-Marquardt			
	K	a	b	c	K	a	b	c
Mean	798.0639	0.1461	9.7859	0.7242	733.3777	0.1460	11.6158	0.7700
Standard	7.8984	0.0016	0.0005	0.0000	59.1496	0.0525	1.8232	0.0741
Deviation Lowest	772.2730	0.1426	9.7849	0.7242	578.2892	0.0251	7.8707	0.5724
Highest	811.5245	0.1504	9.7872	0.7242	773.4050	0.2014	14.2153	0.8084
Oliveira (2019)	761.5322	0.1454	9.7878	0.7243	761.5322	0.1454	9.7878	0.7243

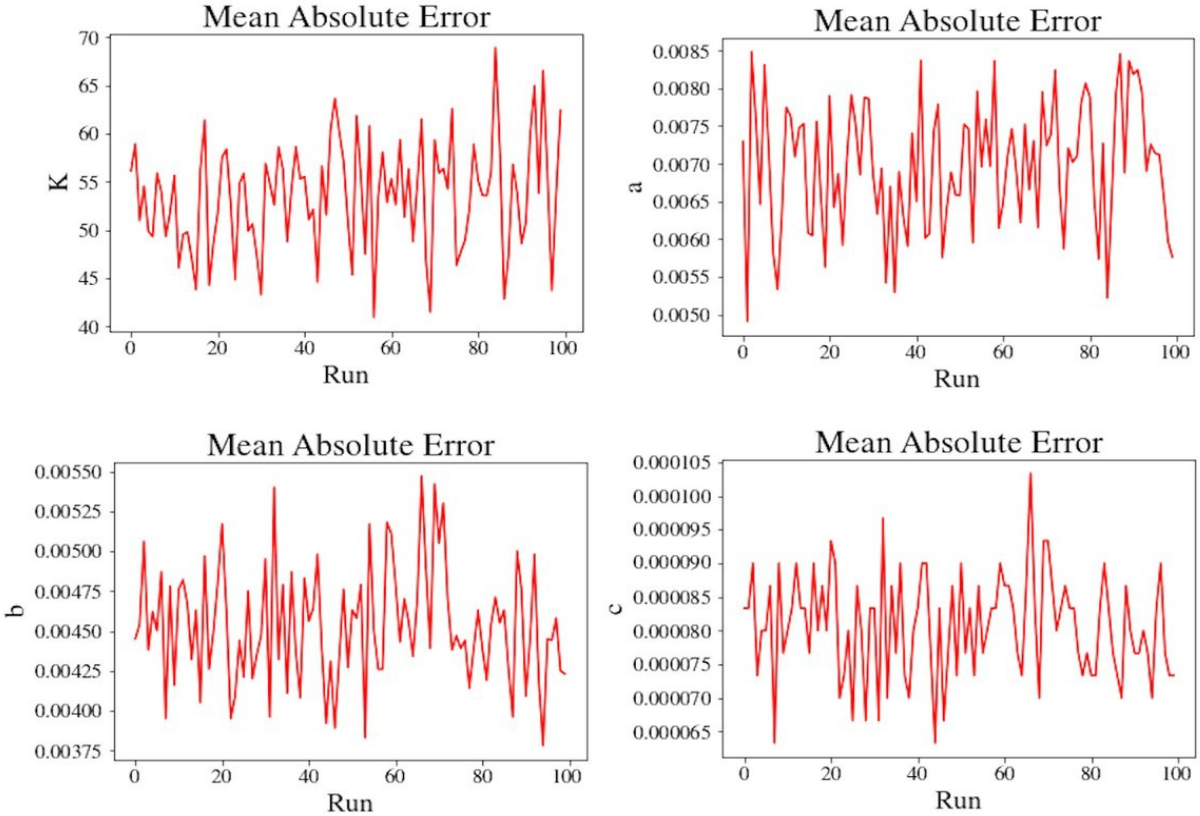

Fig. 2 MAE of ML for K, a, b and c.

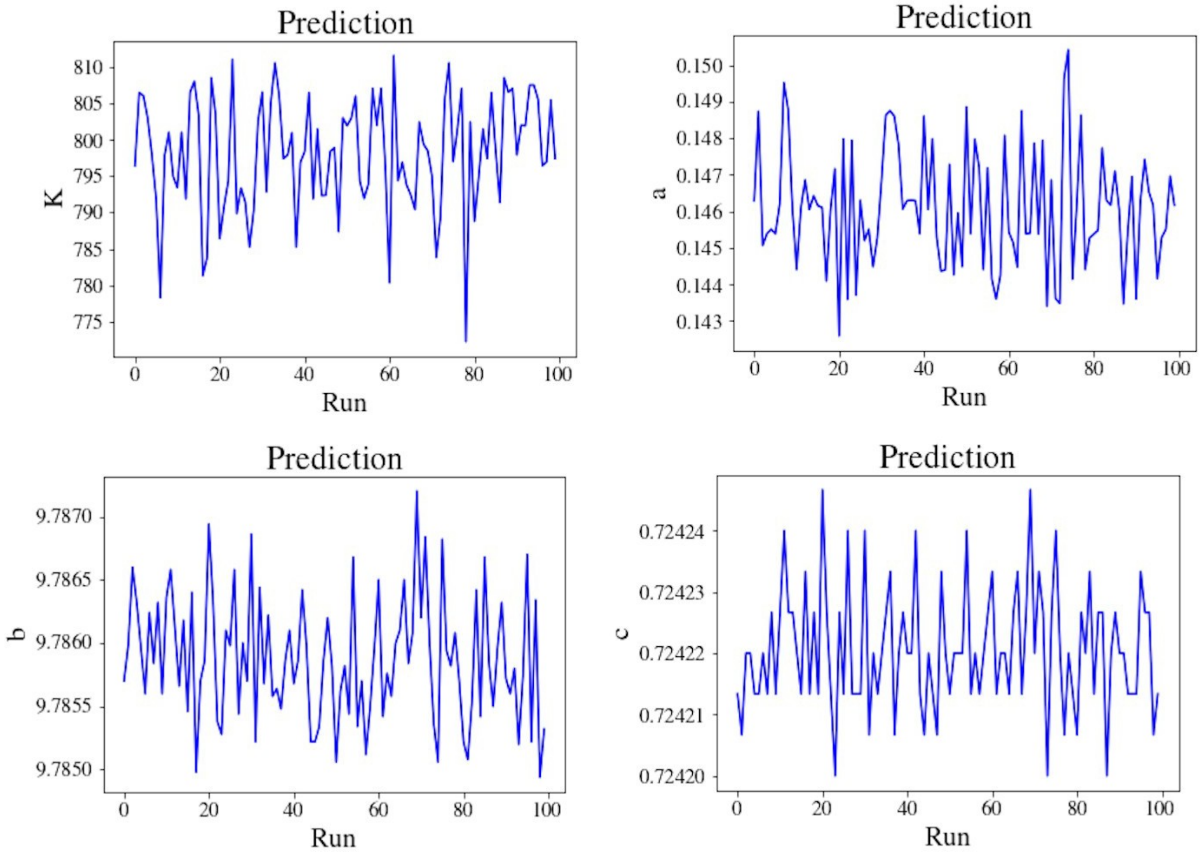

Fig. 3 Prediction of ML for K, a, b and c parameters.

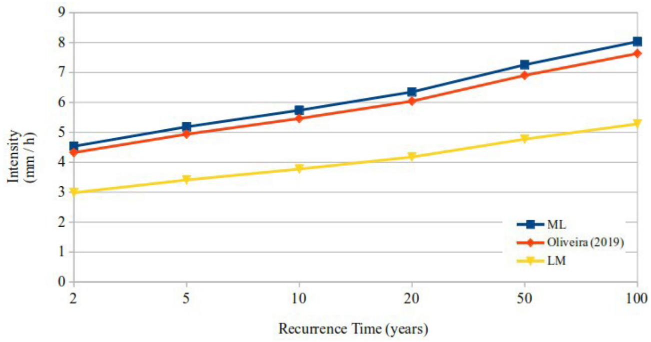

Fig. 4 Intensity of rain

4. Other machine learning techniques can and should be used to try to improve the parameters even more, especially the K and a parameter, which are naturally parameters that have greater variation within the same region.

5. The parameters estimated by the Levenberg-Marquardt method are slightly different due to the pluviometric data being from satellite and having good precision, however, it estimated satisfactory results and compatible with the Varre-Sai region.

REFERENCES

Borrelli, L., Ciurleo, M., Gulla, G. (2018). Correction to: Shallow landslide susceptibility assessment in granitic rocks using GIS-based statistical methods: the contribution of the weathering grade map. Landslides, 15, 1127-1142. https://doi.org/10.1007/s10346-018-0947-7

Botzen, W., Estrada, F., Gooijer, J., Leiser, D., Schoenmaker, D., Siegmann, A., Stork, H., Wagner-Egger, P. (2020). Climate of the middle: Understanding climate change as a common challenge. Ebook. Springer briefs in climate studies: Springer. 76p.

Chang, M., Zhou, Y., Zhou, C., Hales, T.C. (2021). Coseismic landslides induced by the 2018 Mw 6.6 Iburi, Japan, Earthquake: spatial distribution, key factors weight, and susceptibility regionalization. Landslides, 18, 755-772. https://doi.org/10.1007/s10346-020-01522-3

Chang, S. H., Wan, S. A. (2015). Discrete rough set analysis of two different soil-behavior-induced landslides in National Shei-Pa Park. Geoscience Frontiers, 6, 807-816. https://doi.org/10.1016/j.gsf.2013.12.010

Duc-Hung, L., Cong-Kha, P., Trang, N. T. T., Tu, B.T. (2012). Parameter extraction and optimization using Levenberg-Marquardt algorithm. Fourth International Conference on Communications and Eletronics, 434-437. https://doi.org/10.1109/CCE.2012.6315945

Fu, Z., Han, B., Chen, Y. Levenberg-Marquardt method with gerenal convex penalty for nonlinear inverse problems. Jounal of Computational and Applied Mathematics, 404, 2022. https://doi.org/10.1016/j.cam.2021.113771

Hijiao, Y., Xiaohu, W., Min, W., Danrui, S., Jun, W., Ying, Z. Data-based groundwater quality estimation and uncertainty analysis for irrigation agriculture. Agricultural Water Management, 2022. https://doi.org/10.1016/j.agwat.2021.107423

Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., Ahmad, B. B. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena, 163, 399–413. https://doi.org/10.1016/j.catena.2018.01.005

Huang, F., Cao, Z., Guo, J., Jiang, S.-H., Li, S., Guo, Z. (2020). Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping, Catena,191. http://dx.doi.org/10.1016/j.catena.2020.104580

Hudnurkar, S., Rayavarapu, S. (2022). Binary classification of rainfall time-series using machine learning algorithms. International Journal of Electrical and Computer Engineering, 12(2), 1945-1954. http://doi.org/10.11591/ijece.v12i2.pp1945-1954

Junior, J. L., Silva, L. P., Kalas, F. A., Rodrigues, P. P. G. W., Alvim, J. C. (2019). Reservoir implantation for flood dampening in the Macaé River basin using the Mohid Land model, Ciência e Natura, 41. https://doi.org/10.5902/2179460X31860

Li, L., Qiao, J., Yu, G., Wang, L., Li, H. -Y., Liao C., Zhu, Z. (2022). Interpretable tree-based ensemble model for predicting beach water quality. Water Research, 11. https://doi.org/10.1016/j.watres.2022.118078

Liu, W., Luo, X., Hunang, F., Fu, M. (2017). Uncertainty of the soil–water characteristic curve and its effects on slope seepage and stability analysis under conditions of rainfall using the Markov Chain Monte Carlo method. Water, 9, 758. https://doi.org/10.3390/w9100758

Nketia, K. A., Asabere, S. B., Ramcharan, A., Herbold, S., Erasmi, S., Sauer, D. (2022). Spatio- temporal mapping of soil water storage in a semi-arid landscape of northern Ghana – A multi-tasked ensemble machine-learning approach. Geoderma, 410. https://doi.org/10.1016/j.geoderma.2021.115691

Oliveira, L. F. C. (2019). Chuvas extremas no Brasil: modelos e aplicações. Lavras: Ed. UFLA, 2019. 388p.

Seroro, L., Joo, H. B., Jiyeong, H., Dongseok, Y., Panos, P., Pasquale, B., Jae, E. Yang., Jonggun, Kim., Kyoung, J. L. (2022). Estimation of rainfall erosivity factor in Italy and Switzerland using Bayesian optimization based machine learning models. Catena, 211. https://doi.org/10.1016/j.catena.2021.105957

Silva, A. T., Lugon Junior, J., Telles, W. R. (2021). Equação de chuvas intensas e vazão da cidade de Itaperuna-RJ. Anais do 6º simpósio sobre sistemas sustentáveis, 3, 104-116.

Tavares, L. P. D. S., Junior, J. L., Silva, J. A. F., Wasserman, J. C., Rodrigues, P. P. G. W. (2018). Water management and urban flood mitigation: Studies and proposals for the macaé river basin in Brazil. Journal of Urban and Environmental Engineering, 12(2), 188–200. https://doi.org/10.5902/2179460X31860

Telles, W. R., Rodrigues, P. P. G. W., Neto, A. J. (2016). Automatic Calibration of a Simulator Applied to a Mountain River Employing Experimental Data of Rainfall and Level – Case Study: D'Antas Stream, RJ. Revista Brasileira de Recursos Hídricos, 21, 143-151. http://dx.doi.org/10.21168/rbrh.v21n1.p143-151

Tucci, C. E. M (2015). Hidrologia: Ciência e aplicação. 4ª ed., Porto Alegre: Editora da UFRGS/ABRH. 2015.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G. P. W., Moreno, J. (2015). Optical remote sens ing and the retrieval of terrestrial vegetation bio-geophysical properties - A review. ISPRS Journal of

Photogrammetry and Remote Sensing, 108, 273-290. https://doi.org/10.1016/j.isprsjprs.2015.05.005

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J. J., Rivera, J. P., J., Camps-Valls, G., Moreno, J. (2012). Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. Remote Sensing of Environment, 118, 127-139. https://doi.org/10.1016/j.rse.2011.11.002