XXIV **ENMC**
Encontro Nacional de Modelagem Computacional
XII **ECTM**
Encontro de Ciência e Tecnologia de Materiais

ESTIMATION OF PARAMETERS USED IN THE RAIN INTENSITY-DURATION-FREQUENCY EQUATION WITH AN ARTIFICIAL NEURAL NETWORK

Arthur Azevedo Battisaco[1] – art.battisaco@gmail.com
Alex Tavares Silva[1] – altasilva@gmail.com
Vicente de Paulo Santos de Oliveira[1] – vicentepsoliveira@gmail.com
Jader Lugon Junior[1] – jlugonjr@gmail.com
[1] Instituto Federal Fluminense – RJ, Brazil

Abstract. *The knowledge of the rainfall intensity using the Intensity-Duration-Frequency (IDF) Equation is important for many applications since it can be used to calculate the dimensions of the drainage network and to estimate erosion. This value can be obtained using parameters that vary according to the Region of Interest (RoI), and for an accurate measurement, it is necessary a robust study. This work aims to generate an Artificial Neural Network (ANN) that can be used to estimate the parameters based on RoI coordinates and altitude. The parameters in the Intensity-Duration-Intensity (IDF) equation used to train the ANN were obtained from previous work calculated from pluviometric station public data. Then the same locations were used to compare the results obtained with the trained Artificial Neural Network (ANN) to evaluate its performance against results using the interpolation technique. As expected, the neural network obtained similar results, its generalization ability proves to be valuable for future use in this area of study.*

Keywords: *Hydrological modeling, Rain modeling, Machine Learning.*

1. INTRODUCTION

The study of rain intensity is a very valuable field of research, primarily for defining a determined rainfall characteristic, knowing it is important to prevent possible urban problems, such as flooding, loss of soil nutrients, and other problems (Cecílio et al, 2009). To acquire this information, the estimation of determining parameters that are needed for a certain region can be derived from measurements from pluviographs. A frequent problem that is faced by researchers and engineers is the lack of information for the specific location of interest. In that case, an estimation can be obtained using interpolation between the closest points locations (Beltrame et al., 1991).

The intensity-duration-frequency (IDF) equation (Eq. 1) is used to calculate the intensity based on the desired return period (T) and duration (t). The adjustment of the parameters K, a, b, and c parameters can be performed using an empirical process, being based only on the pluviograph data for the specific region (Campos et al, 2014).

$$IDF = \frac{KT^a}{(t+b)^c} \tag{1}$$

Where IDF is the average maximum precipitation intensity (mmh^{-1}), T is the Return period (years), t is the Duration of precipitation (min), and K, a, b, c are the IDF parameters relative to the location.

Pruski et al. (2002) developed software (Pluvio 1.3), which aims to assist other researchers to obtain this parameter without the need of measurements and a previous study at the location of interest, using the coordinates. This is performed using an interpolation procedure and the information for thousands of pluviometric stations across Brazil. The interpolation is independent for each one of the parameters and uses the measured information from the State that the user needs.

Usually, the interpolation method is good enough, mainly using the ponderation factor as the inverse square of the distance between the desired location which one wants to know the group of parameters, and the location where the parameter was already established. The method obtained a mean error of 18,65% to 19,83%. (Cecílio & Pruski, 2003).

The interpolation is performed using only the latitude and longitude coordinates (X and Y coordinates), ignoring the altitude at the locations of interest and at the locations for which the parameters are already known. The authors didn’t find articles studying the influence of altitude in the IDF equation parameters.

One important question in our research is whether adding altitude to the model would influence the estimation of the parameters so that it would result in better results compared to a model that does not use it. For this, we used an Artificial Neural Network, mainly because of its capacity to receive the use of multiple-input values.

2. METHODOLOGY

2.1 The data coordinates and parameters data

The data used was taken from the PLUVIO 2.1 software (Silva et al., 1999), which by itself can give the parameters from ROI's where there is no pluviometric station. In fact, the database for the areas that were already known was very important to obtain. So each station was labeled using the latitude and longitude at our disposal, including the parameters a, b, c, and K. The data was organized in such a way that it would be possible to use them to train the Artificial Neural Network (ANN).

After the coordinates were obtained, it was necessary to transform the unit that was used on Pluvio 2.1 (Degrees, minutes, and seconds) to one that would be more understandable to the ANN, so it was converted to decimal degrees. Another procedure was to obtain the height for every pair of coordinates, the majority of the coordinates we got from the site of Instituto Nacional de Meteorologia (INMET) which has a list for all the pluviometric stations in Brazil.

For the data, previous analysis histograms were plotted, in a way to ascertain any existence of outliers or any type of anomalies such as any mislabelling problems. To inquire about any type of correlation between the parameters and the coordinates, correlation graphics were plotted and analyzed.

When the analysis was performed, it was found that it was not necessary to implement an ANN to model the parameter K, because its correlation with the C parameter was very high. So, using a simple linear regression one would be able to arrive at satisfactory results for it. However, this logic does not apply to the other parameters, so it was necessary to apply the training process to finally have a model that can predict them.

2.2 The Artificial Neural Network design

After that analysis, it was concluded that it would be better to develop one ANN for each parameter. In this case, it's in the complex modeling field, creating one model that would predict the four parameters would require a robust system and a deep neural network. So to solve this problem, it would be better to estimate each one at a time and consequently reducing the error rate and processing time.

Each of the ANN was based on basic neural network architecture, where it's constituted of neuron and fully connected layers. Similar to the work realized by Battisaco et al. (2020). There was no need for a complex model, although the data is a case of complex modeling, a simple shallow ANN is capable of modeling such complex behavior, such that errors between actual and predicted results are often acceptable.

Using the grid search method, the best hiper parameters were searched for each ANN. The process to find these parameters is rather slow, because, to acquire the optimal states, requires many iterations, and even automating the process would need human input. This step of training is still an empirical process (Hegazy et al, 2008).

As mentioned earlier, our data is limited, so we limited the area of prediction of our ANN to the southeast region of Brazil, for which there is more data collected from the pluviometric stations, and consequently, more parameters are known. Other regions suffer from this lack of data, mainly the north and midwest, which means it could be a potential problem for future research in these areas, yet PLUVIO has results for these regions from its interpolation model.

2.3 Checking the results

After training the ANN, it obtained the first set of values for the parameters a, b, c, and K from pluvio. Then it was possible to validate the results using estimations performed by the PLUVIO 2.1 software for random locations within the area of study (Brazilian southeast).

3. RESULTS AND DISCUSSION

Almost every data was gathered from the Pluvio database, which in total correspond to 373 unique pluviometric stations across the southeast region. The study was limited to this location for two main reasons, the first it's important to understand that some areas have pluviometric patterns more similar than others, so limiting the region leads to a possible improvement in predictions by not running away from their rainy behavior, and the second by the lack of data as mentioned before.

To implement the study, it was necessary to know the height for the points of interest. One of the main assumptions is that height can improve the performance of models to predict this type of parameter, since until then those presented to us only used latitude and longitude, ignoring the altitude factor, which directly affects the temperature and humidity of a particular location.

One way to validate this assumption was to verify some kind of correlation between the parameters and the input variables (altitude, longitude, and altitude), although there may not be a direct correlation, it may occur when two of these variables or even the three are used in the system.

3.1 *Data correlation and the Artificial Neural Network*

The first step was to perform the study of the correlation between each input variable and the output parameter so that it would be possible to observe if there was any direct relationship between them. Besides the correlation chart, the histogram chart was also plotted so that one could see the existence of an outlier more broadly.

The most noticeable result is shown in Fig. 1, where is presented the histogram and the correlation graphs for parameter a. From the graph it is possible to see that there is no direct linear relationship between the parameters, except with the length, where the lower it is, the higher the value of a, yet its Pearson correlation coefficient is only 0.036, i.e., a very low correlation (Schober et al., 2018). This result means that there is no direct influence on the parameter A, besides this value does not have many outliers and presents a normal distribution

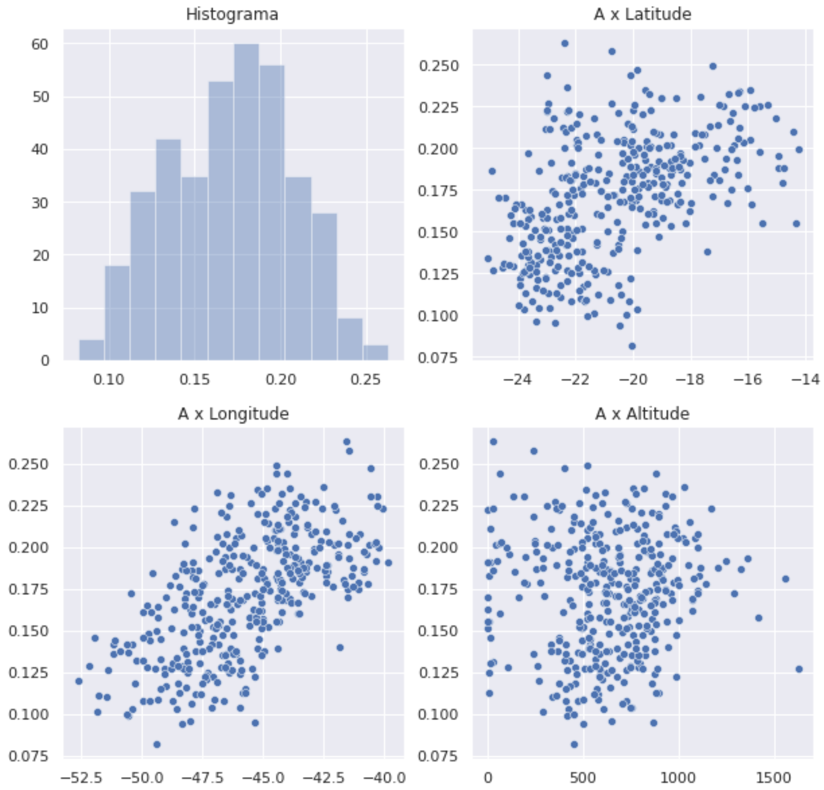

Figure 1 - Parameter a histogram and scatter plot considering Latitude, Longitude, and altitude

The same process was also applied to the other parameters b, c, and K. None of them showed any significant correlation to be mentioned and their histograms did not present anything out of the ordinary, all of them with a normal distribution as expected.

To determine the need for an Artificial Neural Network for each parameter, it was studied the correlation graphs between the variables. Only two variables were correlated, K and c, and with a Pearson correlation coefficient of 0.86, which is an indicator of a strong relationship between them. So instead of generating an ANN to predict parameter c, a polynomial regression was built between these parameters. Figure 2 shows the graph of K and c.

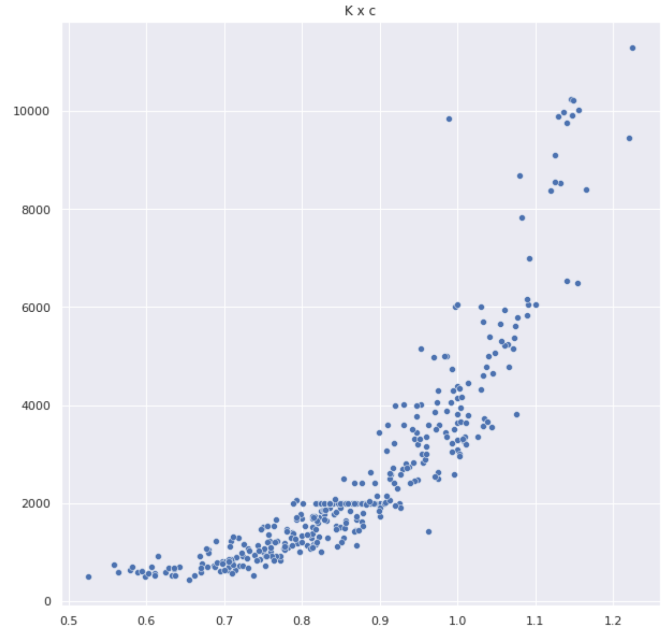

Figure 2 -Graph for the correlation between parameters K and c.

3.2 Result explanation

The hyperparameters that were used for each ANN are shown in Table 1.

Table 1 - Hyperparameters for each ANN trained

Parameter	Number of Hidden Layers	Number of neurons per layer	Number of epochs	Activation function	Learning rate
a	4	36/36/24/24	2,500	ReLu	0.001
b	4	30/30/30/30	3,000	ReLu	0.001
K	4	36/36/18/18	2,000	ReLu	0.001

It is worth mentioning that all input and output data were normalized before the training process so that the algorithm could adapt better to them. This procedure results in a significant reduction in both the error and the training time (Sola, J., Sevilla, J. 1997). In general, the mean square error of the predictions was below 0.09. The one with the smallest error was parameter a, with 0.03. This shows that in a way the models were able to adapt well to this data set.

Comparison of results between the ANN and interpolation

To check the performance of the Artificial Neural Networks, it was taken seven random stations across the southeast, cataloged their latitude, longitude, and altitude in such a way that they could be used as input for the ANN. Then, the results obtained with the ANN were compared to parameters measured by stations. All the results are shown in Table 2.

Table 2 - Results obtained using the ANN and measured by stations

Neural network results				Measured points			
A	B	C	K	A	B	C	K
0.219	44.887	1.136	8665.939	0.205	40.691	1.165	8394.510
0.232	30.325	0.858	2568.669	0.225	29.391	0.944	2450.458
0.265	20.110	0.777	1397.272	0.258	19.294	0.855	1497.781
0.123	24.628	0.844	1770.490	0.126	25.000	0.829	1844.723
0.134	20.818	0.855	1432.783	0.129	19.720	0.870	1662.659
0.236	26.712	0.847	3148.447	0.236	24.664	0.975	2629.477
0.190	23.298	0.867	1910.414	0.182	21.410	0.767	1652.972

In Table 3 are presented the percentage errors of the estimates, in that case, considering Pluvio 2.1 as a reference.

Table 3 – Absolute percentage error for the estimates obtained with the trained ANN

Percentage error			
A	B	C	K
6.9%	10.3%	2.5%	3.2%
3.2%	3.2%	9.1%	4.8%
2.6%	4.2%	9.1%	6.7%
2.1%	1.5%	1.8%	4.0%
3.5%	5.6%	1.7%	13.8%
0.1%	8.3%	13.1%	19.7%
4.6%	8.8%	13.1%	15.6%

The average absolute percentage error for this point was 3.3% to a, 6% to b, 7.2% to c, and 9.7% to K. The largest percentage error is that of K, this much due to its highly complex behavior, showing no direct relationship with the input variables. It is interesting to note that even though K had a high error, C, which was a predicted parameter relative to K, did not have the same error. The parameters a and b had better errors than we expected, showing that it is feasible to use an ANN to model this complex function.

For a final comparison, the parameters were used in the IDF equation to see how the new results look like, using the same parameters that were shown in table 3, we calculated the IDF function using different T and t, the results are on the table below:

Table 4 - Percentage error for the IDF equation results

		T values		
		20 years	30 years	60 years
t values	30 min.	Average: -12%	Average: -13%	Average: -13%
	60 min.	Average: -13%	Average: -14%	Average: -14%
	90 min.	Average: -14%	Average: -14%	Average: -15%

The absolute error analysis is not so different from the analysis of the parameter, whoever, it is noticeable that the results are percentually far apart. The T parameter don't appear to contribute too much with the error, in comparison with the t, which in the formula is raised by the parameter a and summed with b. In general, the T which is the return time in years isn't is as impactful than t even using the measured parameters.

Other thing to take account it that if one of the parameters has been predicted with a high error rate, this error is carried to the final result in the IDF function, and considering that parameter b had a high error, it is to be expected that the results also follow this behavior

4. CONCLUSIONS

It can be inferred that the model is not bad, but it is still not reliable enough to be used as a source of information for these parameters like Pluvio 2.1, even so, we see that the model can be improved, especially taking into account that the efficiency of a neural network is strongly related to the quantity and quality of training data, which was scarce during the preparation of this work.

The use of altitude proved to be a promising addition for future models not only such as neural networks but also other mathematical modeling tools. Although possibly at first glance the one that best suits three different input variables would be ANN algorithms, other algorithms within the machine learning group are to be tested in the future.

The advantage of stimulating the field of forecasting heavy rains with the help of artificial intelligence brings a beneficial union between technology and environmental management. These technological advances can bring benefits to prevent environmental catastrophes or even help in areas such as agriculture where there are problems caused by the lack or excess of rain.

Acknowledgments

The authors acknowledge the financial support provided by CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Finance Code 001), CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, and FAPERJ, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro.

5. REFERENCES

Battisaco, A.A.; Lugon, J.J.; Paiva, P.M. (2020), “Uso de neural artificial para estimativa da altura de pluma de sedimentos gerada durante a remoção de dutos flexíveis offshore”, XXIII Encontro Nacional de Modelagem Computacional, 2020, 529-537

Beltrame, L.F.S.; Lanna, A.E.L.; Louzada, J.A.S (1991), Chuvas intensas. Porto Alegre: IPHUFRGS,69

Campos, Alcinei Ribeiro et al (2014), Equações de intensidade-duração-frequência de chuvas para o estado do Piauí. Revista Ciência Agronômica., v. 45, n. 3, pp. 488-498.

CECÍLIO, R. A. et al (2009), Avaliação de interpoladores para os parâmetros das equações de chuvas intensa no Espirito Santo. Ambi-Água, v. 4, n. 3, p. 82-92.

Damé, Rita de C. F., Teixeira, Claudia F. A. e Terra, Viviane S. S (2008).Comparação de diferentes metodologias para estimativa de curvas intensidade-duração-freqüência para Pelotas - RS. Engenharia Agrícola. v. 28, n. 2, 245-255.

Hegazy T., Moselhi O., Fazio P. (1994), “Developing practical neural network applications using backpropagation.”, Microcomputers in Civil Engineering, 9(2),145–59.

Pruski, F.F.; Silva, D.D.; Teixeira, A.F.; Silva, J.M.A.; Cecílio, R.A.; Silva, D.F (2002). “Plúvio 1.3: chuvas intensas para o Brasil”.Congresso Brasileiro de Engenharia Agrícola, 31, 2002, Salvador. Anais... Salvador: Sociedade Brasileira de Engenharia Agrícola.

Cecílio, R.A. e Pruski, F.F (2003).Interpolação dos parâmetros da equação de chuvas intensas com uso do inverso de potências da distância. Revista Brasileira de Engenharia Agrícola e Ambiental. v. 7, n. 3, pp. 501-504.

Silva, D.D. et al (1999). Estimativa e espacialização dos parâmetros da equação de intensidade-duração--frequência para o Estado de São Paulo. Engenharia na Agricultura, Viçosa, v.7, n.2, 70-8.

Sola, J., Sevilla, J. (1997), “Importance of input data normalization for the application of neural networks to complex industrial problems”, IEEE Transactions on Nuclear Science, 44(3), 1464–1468.

Schober, P., Boer, C., Schwarte, L. A. (2018), “Correlation Coefficients.”, Anesthesia & Analgesia, 126(5), 1763– 1768.