



# Use of Spatial Visualization for Pattern Discovery in Evapotranspiration Estimation

Uso de Visualização Espacial para Descoberta de Padrão na Estimativa de Evapotranspiração

*Fernando Xavier*<sup>1</sup>  
*Maria Luíza Correa Brochado*<sup>2</sup>

Recebido em abril de 2018.  
Aprovado em novembro de 2018.

## ABSTRACT

In Water Resources area, data are obtained from various sources, such as measuring instruments and satellites. Often such data may contain patterns that are not easily identified, either because of the large volume of data sets or because the analysis requires the use of several data dimensions. In this way, this study proposes the application of machine learning resources and spatial visualization to identify patterns in the estimation of an important component of the hydrological cycle: the evapotranspiration. This work is expected to contribute to an approach to estimate evapotranspiration, using spatial resources for pattern identification and model generation.

**KEYWORDS:** Evapotranspiration. Spatial Visualization. Machine Learning. Water Resources.

## RESUMO

Na área de Recursos Hídricos, dados são obtidos através de diversas fontes, como instrumentos de medição e satélites. Muitas vezes, esses dados podem conter padrões que não são facilmente identificados, seja pelo grande volume de conjuntos de dados ou porque a análise demanda o uso de diversas dimensões de dados. Dessa forma, este estudo propõe a aplicação de recursos de aprendizado de máquina e visualização espacial para identificação de padrões na estimativa de um importante componente do ciclo hidrológico: a evapotranspiração. Espera-se com este trabalho

---

<sup>1</sup> Graduate Program in Electrical Engineering - Polytechnic School - University of São Paulo (USP), Brasil. E-mail: fxavier@usp.br

<sup>2</sup> Graduate Program in Geography - Geography Department - University of Brasília (UnB), Brasil. E-mail: luizacorreaenf@gmail.com

contribuir com uma abordagem para estimativa da evapotranspiração, usando-se recursos espaciais para identificação de padrões e geração de modelos.

PALAVRAS-CHAVE: Evapotranspiração. Visualização Espacial. Aprendizado de Máquina. Recursos Hídricos.

\* \* \*

## Introduction

In Big Data scenario, a big challenge for researchers is how to extract useful information in large volumes of data which are obtained from various sources and at increasing rates. According to van der Aalst (2014), more than to collect more data, the goal is to turn these data into real value. This context is reinforced by the fact that the data are generated / obtained in various ways and means, such as social networks and devices related to the Internet of Things, in several areas of knowledge (VAN DER AALST, 2014).

In Climate Change area, for example, data are generated by mathematical models, sensors, remote sensing, conventional measurement instruments (AGARWAL AND BISWAS, 2017). Processing this data and extracting information in this context requires the use of new approaches or even the adaptation of existing approaches, with application of many techniques and concepts in an integrated way.

Among these approaches, stands out the machine learning, by use of techniques in which data is processed in an automated way using algorithms with a variety of objectives, such as patterns discovery in large volumes of data. Machine learning systems have been widely used in many fields that uses complex datasets, such as astronomy, biology, climatology, medicine, finance and economy (AL-JARRAH et al., 2015).

Another widely used technique is spatial visualization, in which georeferenced data are analyzed combining spatial layers. Spatial analysis provides many visualization benefits in areas of high data density but, in other hand, occurs frequently overlapping that makes it difficult to distinguish between the points. One possibility that solves this problem is the

mapping by heat maps, technique which uses a color gradient to represent the geographic density of elements on a map.

The use of these techniques, applied alone or together, may reveal patterns that are not clear from the preliminary analyzes and should be used to confirm or refute hypotheses about the data, adding new information to the experiments. Moreover, spatial visualization for Big Data analysis can help analysts and decision makers obtain insights more quickly (PFEIFFER and STEVENS, 2015).

Based on this, this article is an application report of an experiment using two of these techniques, data mining and spatial visualization, for the discovery of regional patterns in an important activity in the Water Resources area: the evapotranspiration estimation.

Using data available from the Brazilian National Institute of Meteorology (INMET), it was applied the data mining technique to identify a model for estimating evapotranspiration in a meteorological station, named reference station. This model, in turn, was applied to other meteorological stations in Brazil, aiming to identify the validity of this model in data for these stations. Finally, the data generated were visualized on maps in order to verify if there was a regional pattern, be it related to latitude or vegetation cover.

It is expected, with this study, to demonstrate how the integrated use of two techniques for data analysis can aggregate information that would not be clear if they were applied in an isolated way. In addition, the experiment may reveal information to researchers in the Water Resources area about possible patterns in the evapotranspiration estimation.

In Section 2, it is described the theoretical reference used in this work, related to evapotranspiration and its estimation methods. The following section describes the approach used in this work and, in Section 4, it was detailed the steps of the experiment. The second experiment is described in the Section 5, followed by the analysis about results obtained in the Section

6. Finally, in Section 7, final considerations of this work are made, including suggestions of future research works.

This paper is an extended version of XAVIER e BROCHADO (2017), presented in XVIII Brazilian Symposium on GeoInformatics (GEOINFO 2017).

## **2 Background**

### **2.1 Evapotranspiration**

Evapotranspiration is a component of the water cycle, defined as the loss of surface water through the combination of soil evaporation processes and vegetation transpiration (FAO, 2017), which returns to atmosphere as vapor.

There are many local and meteorological factors that affect the amount of the water lost by the surface in the evapotranspiration process (FAO, 2017). Among the local factors, are included the type of vegetation and the soil, which influence the surface capacity to absorb the water received from rain. In addition, weather conditions such as temperature, wind speed, humidity and cloudiness also contribute to the process, directly or indirectly, affecting the energy amount used in the transformation of water molecules from the liquid to the gaseous state.

According to the FAO, the main energy source used in this process comes from solar radiation, which varies according to factors such as latitude, altitude, cloudiness, period of the year and the time of day.

The estimation of evapotranspiration can be done through instruments, remote sensing methods and mathematical models, such as the Penman-Monteith reference equation, FAO reference method. Due to the difficulty of obtaining the values for all parameters of the equation (MAJIDI et al., 2015), other alternative models have been proposed, such as the Thornwaite and Hargreaves equations (CAMARGO et al., 1999).

According to the FAO Guide, one of the alternatives for evapotranspiration estimation in the missing data scenario would be using data from nearby meteorological stations, since the conditions that affect evapotranspiration may be similar in geographically close regions.

## 2.2 Machine Learning application in Evapotranspiration Estimation

Machine learning has been widely used in forecasting activities using data related to the area of Water Resources. Unlike traditional approaches, machine learning is used in a data-driven approach, in which historical data is processed in order to generate prediction models.

In Rhee and Im (2017) work, it was used machine learning approach to generate a drought forecast model, using precipitation and temperature data from a 42 years data series. Xingjian et al. (2015) used neural networks as an approach to solve the precipitation nowcasting problem, using radar data. They compared their models with a state-of-the-art algorithm and got better results.

Aiming to predict monthly precipitation and evapotranspiration, Deo and Şahin (2015) conducted a study that applied neural networks on data of eight stations in eastern Australia. This study used climate variables over the period 1915–2012 and achieved good results, when compared with historical data. Also using neural networks, Feng et al. (2017) made a study to estimate daily evapotranspiration with only temperature data, generating a more simpler model than other approaches commonly used. Their study was made with data from six stations from 1961 to 2014.

Similar to Feng et al. (2017) work, Xavier (2016) applied a machine learning approach to estimate evapotranspiration using historical data from the National Institute of Meteorology (INMET). In this work, however, it was applied data mining to discover a model by locality, using data from 2010 to 2014. This study aimed to generate a model with less attributes than the Penman-Monteith equation.

Aiming to discover regional patterns, Xavier, Tanaka and Revoredo (2015), developed a study using machine learning to generate evapotranspiration estimation models, based on the hypothesis that similar locations could use the same model to estimate the evapotranspiration. In this study, the model discovered by the machine learning method for one locality was applied in other five locations, using the latitude as a factor of similarity. According to the preliminary results, the model learned for a locality could be applied in similar locations.

These studies have in common the use of machine learning methods to discover alternative estimation models, using historical data and aiming to simplify the traditional approaches. However, how could these approaches be improved for use in regions with wide variety of physical characteristics, such as Brazil?

From this research problem, it is intended to evolve the preliminary study conducted by Xavier, Tanaka and Revoredo (2015), aggregating to the machine learning approach the use of spatial visualization. This modified approach aims to discover possible patterns in evapotranspiration estimation in locations with wide variety characteristics, analyzing other similarity criteria other beyond the latitude used in the reference study. Moreover, this study will apply this approach to more stations than preliminary study mentioned previously.

### **3 Methodology**

#### **3.1 Solution Proposed**

Using historical series of meteorological data obtained from INMET datasets, a model of evapotranspiration estimation for a locality will be obtained through the use of machine learning. This location will be called a reference location and will be used in comparison with other locations.

The model generated for the reference location will be applied to estimate evapotranspiration using other locations data and, after that, the correlation will be calculated between the evapotranspiration value obtained by the model learned and the historical values in the INMET datasets.

In this way, the hypothesis to be verified in this work is: how much more a test location is similar, defined according to latitude or vegetation, to the reference location, better will be the correlation obtained by application of the reference equation in the test location data.

### 3.2 Evaluation

To evaluate the proposed solution, two experiments were made: in the first one, it was applied the evapotranspiration model learned from reference station in other stations, aiming to evaluate a possible correlation between latitude and precision of results, comparing with historical data. In this experiment, the proximity with the reference station was the similarity criteria used. In the second experiment, it was used biome as similarity criteria, with one reference station for each Brazilian biome.

By means of spatial visualization, the correlation values of each location will be compared using layers of latitude data, potential evapotranspiration, and state boundaries.

### 3.3 Experiment Planning

According to Ward, Grinstein and Keim (2010), the visualization process consists of several steps, from raw data collection to visualization by users. By means of this process, five stages were defined to execute the experiment proposed in this research work:

- a) Data Collection and Analysis: meteorological data will be collected from the INMET database and will be applied exploratory evaluation of these datasets;

- b) Pre-Processing: application of transformations in data as well as preparation of datasets for processing, excluding non-relevant columns for this study;
- c) Processing: use of classification algorithms, to determine an estimation model of the evapotranspiration for the reference station and application of this model to data of the other stations;
- d) Spatial Visualization: visualization of correlation values in many maps combined to the layers that represent the similarity factors;
- e) Results Analysis: evaluation if the results confirm or refute the hypothesis defined for the experiment.

In addition to the data visualization pipeline, the steps in this experiment were defined according to the steps of the Knowledge Discovery in Databases (KDD) process (FAYYAD, PIATETSKY-SHAPIRO SMYTH, 1996).

## **4 First Experiment**

### 4.1 Description

In this experiment, the goal was to evaluate if latitude could be a good similarity criterion, in which a evapotranspiration model from one station should be used in other stations of similar latitude.

### 4.2 Collection and Analysis Data Step

The data were collected from the Meteorological Database for Teaching and Research, a tool created by INMET to provide historical data to researchers. The following filters were used to collect the data:

- a) Period: from 01/01/1996 to 31/12/2016;
- b) Measurements: All;
- c) Stations: All.



Performing the search using the three filters described above resulted in data from 265 stations. To automate the process, both the search and the download of the data were done through a shell script developed for that purpose, available in <https://github.com/professorxavier/datascience/>.

These datasets are available in CSV format (comma separated values) and, in addition to the historical data series, contains information about the station, such as latitude, longitude and altitude.

From the analysis of the datasets, it was found that some attributes could be removed, due to the high degree of missing data or to contain values that would not influence in an equation, such as date reference and station code. In this way, it was defined that the attributes from the dataset used in the processing step would be: average wind speed, average max wind speed, evapotranspiration potential, total insolation, average cloudiness, total rainfall, maximum average temperature, mean compensated temperature, average minimum temperature and relative humidity average.

These attributes were obtained according to standards defined by the World Meteorological Organization (WMO, 2014).

#### 4.3 Pre-processing Step

To use the data in the processing step, it was necessary a way to extract the metadata information from the historical data series. In this way, it was developed a simple Java program to extract data according to its use in the later steps: stations information and historical series data.

As a result of this step, files were generated in the CSV format (comma separated values) with the data of the historical series as well as a CSV file with the data of the stations.

#### 4.4 Processing Step

In order to learn the baseline model of evapotranspiration estimation, were used data from the Resende-RJ station, due to its proximity to many stations in Brazil, which would guarantee a good number of stations considered similar, using latitude and vegetation criteria.

Data mining activity was executed using the Weka software (FRANK, HALL e WITTEN, 2016), and M5P algorithm on the Resende-RJ station data, generating a model for evapotranspiration estimation. The M5P algorithm was the one chosen for this task, due to its use in the Xavier, Tanaka and Revoredo experiment (2015), making more confident the comparison between results from the previous work with the work presented here.

This generated model has a correlation coefficient of 0.9715 for the Resende-RJ station data. After that, the generated model was applied to data of each station, obtaining correlation coefficient and storing it in a data file to be used in the next steps.

At the end of the processing stage, data were obtained for 252 stations from the model generated for the station of Resende-RJ. These data were recorded in CSV files for use in the later stage of this study.

#### 4.5 Spatial Visualization

From the correlation coefficients obtained for each station in the previous step, it was defined to generate maps to evaluate the results from a spatial perspective. For these maps, the correlation coefficients obtained was classified according to the range of values defined in Table 1.

The tool chosen for the spatial visualization was QGIS, a free software to support spatial analysis and with support for multiple data formats, that was used to prepare the maps (QGIS, 2017).

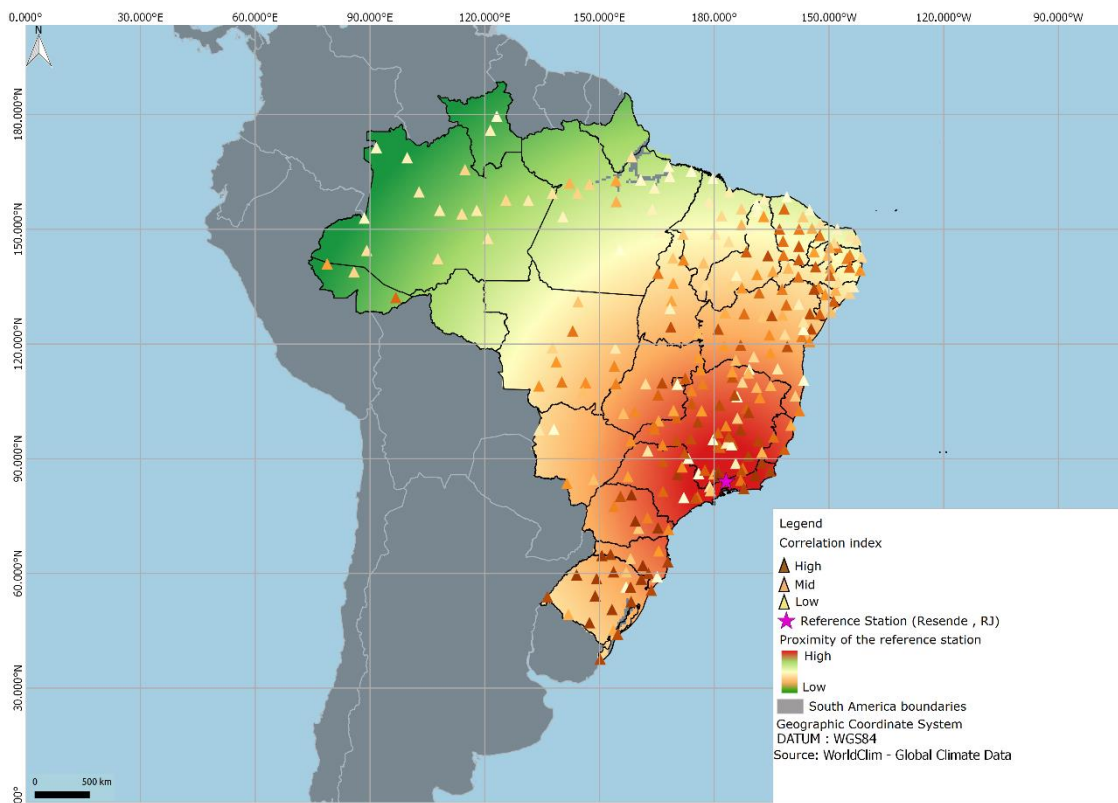
Table 1 - Classification of the correlation coefficient by range

Correlation Coefficient	Classification
> 0.70	High
0.40 – 0.70	Medium
< 0.4	Low

Source: Prepared by the authors

In order to verify the quality of the correlation coefficients in relation to the geographic distance of the reference station (Resende-RJ), a map was created using the distance to the reference station (Figure 1). For this analysis, distance bands were established in relation to the reference station and its influence on the correlation coefficients was evaluated.

Figure 1 - Visualization of the correlation coefficients in relation to the reference station



Source: Prepared by the authors

## **5. Second experiment**

### 5.1 Description

For this experiment, it was used the biome as similarity criteria, with application of the evapotranspiration model of a reference station in data from other stations of the same biome.

### 5.2 Collection and Pre-processing Steps

It was used the same data of the first experiment, collected from the INMET database. No pre-processing additional tasks was necessary, because it was used the same pre-processed data from the first experiment.

### 5.3 Processing Step

In this experiment, it was used one reference station for each Brazilian biome (MMA, 2018). The choice of the reference station was made according to the proximity of other stations, defining a location near to the biome middle point. The only exception was with the Mata Atlântica biome, due to its long territory. In this case, it was chosen Natal as reference station, because this station has a fewer missing data degree, when comparing with other stations in the same biome.

After reference stations definition, it was applied data mining to generate an evapotranspiration model for each station, using the Weka software. The Table 2 shows the data for each reference station.

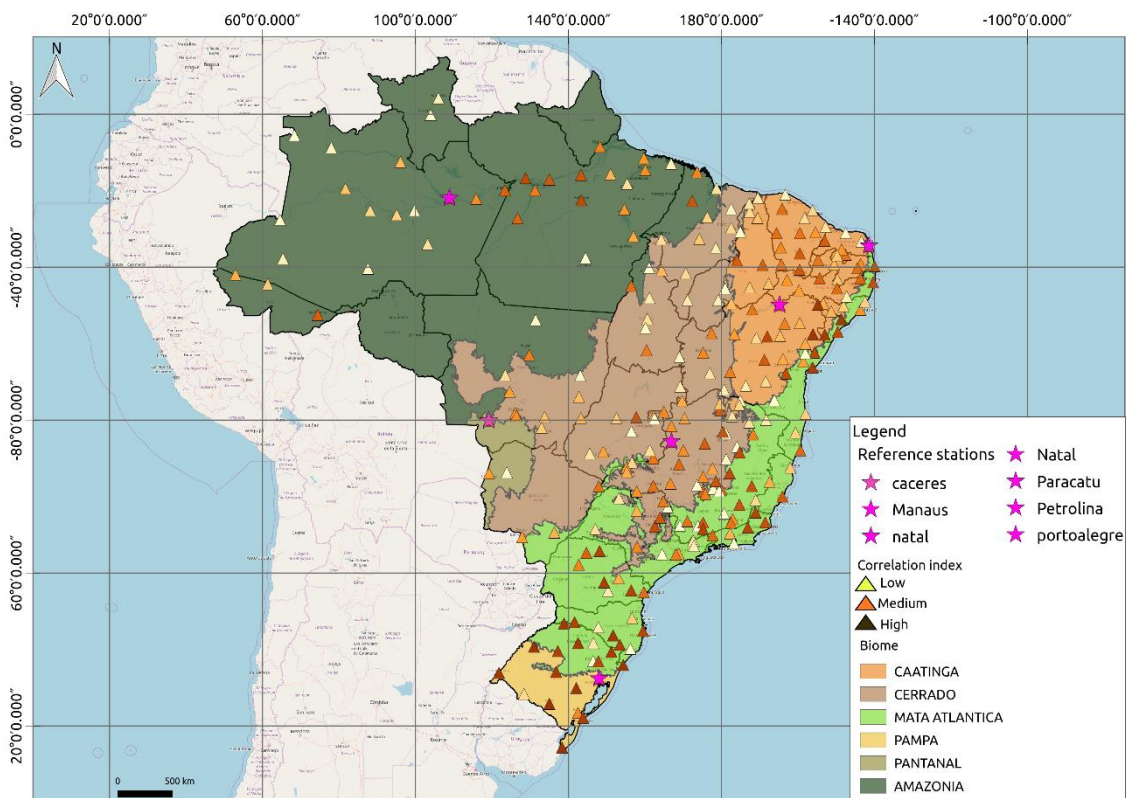
Table 2. Reference stations

Biome	Reference Station	Correlation
Amazônia	Manaus-AM	0.9772
Caatinga	Petrolina-PE	0.9755
Cerrado	Paracatu-MG	0.9678
Mata Atlântica	Natal-RN	0.9409
Pampa	Porto Alegre-RS	0.9937
Pantanal	Cáceres-MT	0.9745

Source: Prepared by the authors

Using a developed java application, each evapotranspiration model was applied in data from all stations of its biome, collecting the correlation coefficient with historical data, as same as was made in the first experiment. This data was stored in a CSV file, for use in further tasks.

Fig. 2- Visualization of the correlation coefficients for Experiment 2



Source: Prepared by the authors

#### 5.4 Spatial visualization

In order to discover possible patterns, as in Experiment 1, the results of Experiment 2 were shown on maps. Although the quantitative analysis showed that there was improvement in the results, as described in section 6.2, the spatial visualization could reveal other results, through of multiple views and layers.

In Figure 2, the classification (High, Medium or Low) of the correlation coefficient for each station was shown on the map, indicating which biome belongs as well as the reference station for each biome. The goal with this view was to check if there was any pattern in the results according to biomes, which would not be easily verified in the raw data.

It was possible to observe that, in the Atlantic Forest, Caatinga and Pampa, there was concentration of medium-high results, with few points classified as “Low”. In the Mata Atlântica biome, it was noticed that the majority Low-points was found in São Paulo and Minas Gerais states.

In Caatinga biome, there are some locations classified as Medium category and the majority of them are located in locations near to ocean. It was noticed too that some Medium-points are located biomes limits, that could suggest a wrong categorization of these locations.

The most of Medium-points is located in Cerrado biome, that was the biome with worst result for this approach. Comparing with first experiment, there wasn't a great improvement using the approach biome-based.

For the Amazonia biome, however, there was a noticeable improvement in the results of the second experiment. This can be explained by the fact that the Amazon is a biome with a set of characteristics of its own, difficult to compare with other biomes.

## 6. Results analysis

### 6.1 First Experiment

The map in Figure 1 indicates that there was a higher concentration of high correlation coefficients in the South, Southeast and Northeast Brazilian regions. This result is an evidence that the quality of the model generated from the data of the station of Resende-RJ is related to common characteristics of these regions, such as climate and vegetation.

It is also noticed that there was concentration of low correlation coefficients in the Amazon region, fact that could reinforce the relationship of the model with the vegetation. Another indication of the model's relation to vegetation can be evidenced by the model quality for the Cerrado and Caatinga biomes (with higher concentration of medium and high correlation coefficients) that are similar in relation to the type of vegetation (adapted to climates drier), low levels of rainfall and less drained soils.

After map analysis, it was noticed that latitude would not be a good indicator of similarity between stations. On the other hand, vegetation could be a better indicator, given the concentration of high correlation indexes in the regions of the Atlantic Forest and Pampas, while in the Cerrado and Amazon, there were concentrations of medium and low correction index, respectively.

### 6.2 Second Experiment

When comparing with the first experiment results, the use of a reference station for each biome improve the results, with more stations classified as good correlation. As show in Table 3, 36 new stations were classified as good correlation.

Table 3 - Experiments comparison

<b>Correlation</b>	<b>Experiment 1</b>	<b>Experiment 2</b>
High	156	194
Medium	56	60
Low	53	11

Source: Prepared by the authors

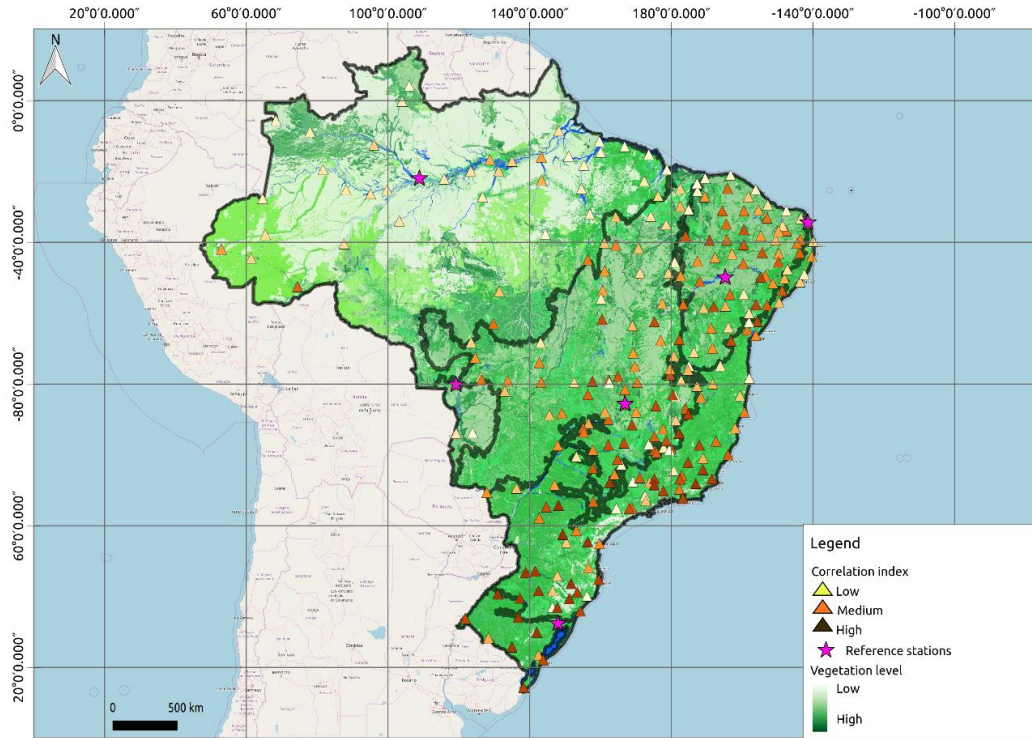
In the Experiment 1, the average correlation was 0.663, while in the Experiment 2, was 0.767. The most observed change was in the Low and High correlation categories, with increasing High Correlation stations in Experiment 2 and decreasing Low Correlation stations in the same experiment.

This information could evidence that biome is a more suitable similarity criteria than criteria only latitude based. A reason for this is that locations in a same biome have many common characteristics, such as vegetation, an important factor for the evapotranspiration process. Even in biomes where there was little improvement, such as the Cerrado, this approach could be used, perhaps with a better reference station choice.

In addition, it was attempted to generate a new spatial visualization to compare the two experiments, adding layer with vegetation data (shown in Figures 3 and 4), for both the data of experiment 1 and experiment 2. It was not observed, besides the differences already noted, a vegetation-related difference for this new approach used in the experiment 2.



Figure 3 – Vegetation layer and correlation coefficients for experiment 1.



Source: Prepared by the authors

Figure 4 - Vegetation layer and correlation coefficients for experiment 2.



Source: Prepared by the authors

## 7. Final Considerations

In the experiment described in this work, it was intending to study the use of spatial visualization as an approach to identify patterns in the estimation of evapotranspiration. Using a machine learning approach, a model was generated for a reference station, which was used to estimate evapotranspiration in other INMET weather stations. The correlation coefficients generated between the evapotranspiration historical data and those calculated by the reference station model were placed on maps to identify possible patterns.

Using latitude and vegetation as factors to classify similarities between stations, it was defined a hypothesis that locations similar to the reference location would have better results with the model generated for this station. The latitude was defined as similarity criteria because it is related to the radiation received by the localities, which is the main energy source in the evapotranspiration process. The vegetation, in turn, was used as criteria mainly due to the fact that part of the evapotranspiration value is originated from the transpiration of the vegetation.

In relation to vegetation, it was observed better rates in the Atlantic Forest biome, which is the same region of the reference station, and in the Pampas biome. In other regions, such as those delimited by the Amazon, the correlation coefficients were low, besides the Cerrado, where the average coefficients were concentrated.

In opposite to the observed in the preliminary study developed by Xavier, Tanaka and Revoredo (2015), latitude was not noticed a good indicator of similarity for estimating evapotranspiration, since there were several points with low correlation coefficients at locations considered similar using this criterion, whereas several points with high correlation coefficients in non-similar locations were identified by the latitude criteria.

In the second experiment, it was noticed an improved result in the correlation coefficient. By use of a specific reference station per biome, more

stations were classified in the group High Correlation. This could suggest that vegetation is a better factor to indicate similarity than latitude, used in the first experiment.

As future works of this study, it is suggested application of the experiment using other reference locations, for test other possible criteria. In addition, due to the variety of factors influencing evapotranspiration, it is also suggested application of the study using the combination of other layers of data, because it is possible that exists other patterns not clearly visible.

A new study could be done, with the aim of improving the results of experiment 2, on the best choice of reference station in each biome. In this work, a midpoint was used, but other criteria could be used.

This research work has demonstrated that the approach of spatial visualization to identify patterns in the estimation of evapotranspiration can be very useful, either to aggregate new information to studies carried out with other approaches or to discover new patterns that were not previously identified in other approaches of analysis of data. The use of spatial visualization, as demonstrated in this study, has brought new perspectives for analyzing data generated by mathematical models, which can be used in other areas of knowledge.

## References

- AGARWAL, P.; BISWAS, S. S. Big Data in Climate Change. **Global Journal for Research Analysis**, vol. 6, n. 7, 2018, pp 50-51.
- AL-JARRAH, O. Y.; YOO, P. D.; MUHAIDAT, S.; KARAGIANNIDIS, G. K.; TAHA, K. Efficient machine learning for big data: A review. **Big Data Research**, vol. 2, n. 3, 2015, pp. 87-93.
- CAMARGO, A. D.; MARIN, F. R.; SENTELHAS, P. C.; & PICINI, A. G. (1999). Ajuste da equação de Thornthwaite para estimar a evapotranspiração potencial em climas áridos e superúmidos, com base na amplitude térmica diária. **Revista Brasileira de Agrometeorologia**, vol. 7, n. 2, 1999, pp. 251-257.

- DEO, R. C.; ŞAHIN, M. Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia. **Atmospheric research**, vol. 161, 2015, pp. 65-81.
- FAO (2017). Introduction to evapotranspiration. Site <<http://www.fao.org/docrep/x0490e/x0490e04.htm>>, accessed in July 2017.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, vol. 17, n. 3, 1996, pp. 37-54.
- FENG, Y.; PENG, Y.; CUI, N.; GONG, D.; ZHANG, K. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. **Computers and Electronics in Agriculture**, vol. 136, 2017, pp. 71-78.
- FRANK, E.; HALL; M.; WITTEN, I. H. The WEKA data mining software: an update. **ACM SIGKDD explorations newsletter**, vol. 11, n. 1, 2009, pp. 10-18.
- MAJIDI, M.; ALIZADEH, A.; VAZIFEDOUST, M.; FARID, A.; AHMADI, T. Analysis of the effect of missing weather data on estimating daily reference evapotranspiration under different climatic conditions. **Water Resources Management**, vol. 29, n. 7, 2015, pp. 2107-2124.
- MMA, Ministry of the Environment. Site <<http://www.mma.gov.br/biomas>>, accessed in June 2018.
- PFEIFFER, D. U.; STEVENS, K. B. Spatial and temporal epidemiological analysis in the Big Data era. **Preventive veterinary medicine**, vol. 122, n. 1-2, 2015, pp. 213-220.
- QGIS, Geographic Information System. Site <<http://qgis.osgeo.org>>, accessed in July 2017.
- RHEE, J; IM, J. Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. **Agricultural and Forest Meteorology**, vol. 237, 2017, pp. 105-122.
- VAN DER AALST, Wil MP. Data scientist: The engineer of the future. **Enterprise interoperability VI**. Springer, Cham, 2014. pp. 13-26.

- XAVIER, F.; BROCHADO, M. L. C. Use of Spatial Visualization for Pattern Discovery in Evapotranspiration Estimation. **Proceedings of the XVIII Brazilian Symposium on GeoInformatics**, Salvador, 2017. pp. 346-356.
- XAVIER, F.; TANAKA, A. K.; REVOREDO, K. C. KDD application on Meteorological Data for Identification of Regional Patterns in Estimation of Evapotranspiration. **Proceedings of the 30th Brazilian Symposium on Databases**, Petrópolis, 2015. pp. 27-32.
- XAVIER, F. **Application of Data Science Techniques in Evapotranspiration Estimation**. Master's dissertation. Federal University of the State of Rio de Janeiro, Graduate Program in Informatics, Rio de Janeiro, 2016. 95p.
- XINGJIAN, S. H. I.; CHEN, Z.; WANG, H.; YEUNG, D. Y.; WONG, W. K.; WOO, W. C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. **Advances in neural information processing systems**. 2015, pp. 802-810.
- WARD, M. O.; GRINSTEIN, G.; KEIM, D. **Interactive data visualization: foundations, techniques, and applications**. AK Peters/CRC Press, 2015. 578p.
- WMO, **Guide to Meteorological Instruments and Methods of Observation**, 6th ed., WMO Rep. 8, World Meteorological Organization, Geneva, Switzerland, 1996.