Geographical Complex Networks applied to describe meteorological data

Aurelienne A. S. Jorge[1], Izabelly C. Costa[1], Leonardo B. L. Santos [2]

[1]Instituto Nacional de Pesquisas Espaciais (INPE)
Cachoeira Paulista, SP – Brazil

[2]Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden)
São José dos Campos, SP – Brazil

aurelienne.jorge@inpe.br, izabelly.costa@inpe.br, santoslbl@gmail.com

Abstract. *Complex Networks have been widely applied to climate data analysis, identifying relations and patterns in the atmosphere on a long-term scale. However, a few investigations have made use of Complex Networks to study meteorology (dealing with short-term changes in the atmosphere). With this in mind, the purpose of the present work is to make some progress in the spatial analysis of metrics in meteorological networks, specifically in precipitation events. We present some results for a study case comprising the Tamanduateí basin, in which we could analyze the spatial dependence intrinsic in the network structure.*

Resumo. *Redes Complexas têm sido largamente aplicadas na análise de dados climáticos, na tentativa de identificar relações e padrões na atmosfera a longo prazo. Algumas poucas pesquisas fizeram uso das Redes Complexas no estudo da Meteorologia (tratando de mudanças a curto prazo na atmosfera). Com base nisso, o presente trabalho tem o propósito de buscar algum avanço na análise espacial de métricas em redes meteorológicas, mais especificamente, em eventos de precipitação. Alguns resultados são apresentados para um estudo de caso compreendendo a bacia do rio Tamanduateí, nos quais foi possível analisar a dependência espacial intrínsica na estrutura da rede.*

1. Introduction

Based on Graph Theory, the study of Complex Networks represents a relevant contribution to science as a tool to describe the structure of a wide range of complex systems in nature and society, such as climate events [Barabási and Pósfai 2016]. In such a context, Complex networks have been applied to climate data analysis, aiming to identify structural patterns and teleconnections. Those researches use similarity measures such as Pearson correlation, event synchronization, or mutual information to construct the network connections. In terms of data, they are based on long time series of atmospheric variables, ranging from months to several years [Tsonis et al. 2006, Boers et al. 2019].

A few works have been held specifically in the weather domain, dealing with short-term changes in the atmosphere and manipulating spatial and temporal high-resolution data through complex networks. One of those few examples handled precipitation data from weather radar, and they achieved significant results in community detection based on a time series of only ten days, with 1km of spatial resolution [Ceron et al. 2019]. The behavior of topological metrics in meteorological networks is a characteristic that remains unknown.

With this in mind, the purpose of the present work is to make some progress in the spatial analysis of metrics in meteorological networks, specifically in precipitation events.

Due to climate changes, extreme precipitation events are becoming more frequent, with several impacts on society. Finding spatial patterns of precipitation events could represent a significant advance in atmospheric science and several applications, from health geography to resilient urban mobility [Santos et al. 2017].

2. Materials and Methods

2.1. Data

The case study presented here was held in São Paulo Metropolitan Region, specifically comprising the area of Tamanduateí basin, from January of 2015. Located on the Tiete river's left margin, the Tamanduatei basin has its source in the city of Mauá. It also crosses the towns of Diadema, São Caetano do Sul, besides the eastern and central zones of São Paulo [Ramalho 2007].

Due to its spatial and temporal high-resolution data, we used weather radar time series as our base dataset. Considering the mentioned study area, the weather radar located in the city of São Roque is the one that offers the best coverage, with a range of 250 kilometers. Its scans provide data with 1 kilometer of spatial resolution every 10 minutes [DECEA 2010]. The raw data is composed of a volume scan which contains scan values for different angles of elevation. For each of these elevation angles, an azimuth scan is performed, and such a scan is called Plan Position Indicator (PPI). The São Roque radar has 15 elevation angles, starting at 0.5 degrees to approximately 20 degrees [Redemet 2015].

For the first study cases, we used PPI data corresponding to the first level of elevation. These data are available in binary files with a grid of values. Such information is reflectivity value (dBZ), which we can convert into estimated rainfall rate. In summary, the higher the reflectivity value, the more intense is the estimated precipitation. As mentioned before, the selected time series comprises the entire month of January of 2015 with a temporal resolution of 10 minutes, so it is composed of more than 4400 scans in time, each one of them including 783 points in space.

2.2. Network Construction and Analysis

Spatial embedding is a physical property inherent in many phenomena modeled through networks, including the meteorological events addressed in this research. Therefore, we use here geographical graphs, which are graphs whose nodes have a known geographic location, and their edges contain an intrinsic spatial dependency [Santos et al. 2017].

We developed a tool to manage the input data and construct the network taking into account its geographical component. We called it GIS4Graph. It delivers output files with topological metrics calculated. One of these outputs is a shapefile, a file compatible with GIS platforms, and allows graph visualization in geographical space.

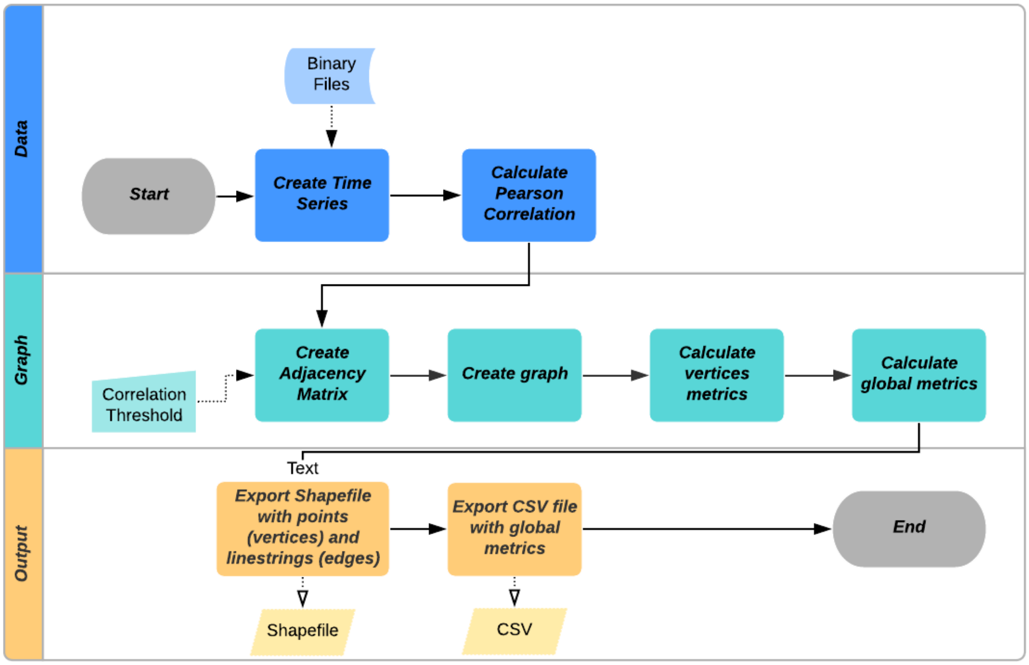

Figure 1. Graph4GIS flowchart

The application is composed of 3 main modules: Data, Graph, and Output. The developed flowchart is presented in Figure 1. The first module is responsible for dealing with the binary files provided by the weather radar. It reads all data and creates a time series for each grid point. Then, it calculates the Pearson correlation between each pair of them.

The second module deals directly with graph construction and metrics calculation. First of all, it generates a weighted adjacency matrix based on the Pearson correlation values - eliminating the ones under a predefined threshold. Next, it builds the graph with one node for each grid point and the edges with the weights indicated by the adjacency matrix. After that, it calculates the topological metrics, both global and nodes specific. Degree, clustering coefficient, and average shortest path are a few examples of them.

The results exportation is done by the third module, which delivers a shapefile with a set of points and lines, geographically representing the graph's nodes and edges. The metrics of each node appears as an attribute of the point in the shapefile. The application also generates a CSV file with all the global network metrics. The Output module also exports some charts to support auxiliary analysis.

We execute the application inputting different correlation thresholds, and we analyze the network diameter in each case. The final network is the one with the highest diameter metric, aiming to promote the best possible balance between removing the least relevant edges and keeping the most important ones - as applied in previous papers in the literature [Santos et al. 2019, Ceron et al. 2019]. The threshold in which the network achieves the highest diameter value is called a critical threshold.

3. Results

Before analyzing the network built for the mentioned case study, we can observe the spatial dependence inherent in such data on Figures 2 and 3. The first one shows how the (temporal) correlation between the (time series associated with each) pairs of points is related to the geographical (euclidean) distance between them. We grouped correlation values into three categories - minimum, medium, and maximum - respectively colored in red, green, and blue. For the red group, it is possible to see a spatial dependence up to approximately 3 km. Regarding medium and maximum categories, the temporal correlation is considerably high between 1 and 10 km of distance, but we can still observe the influence of spatial dependence until 20 km.

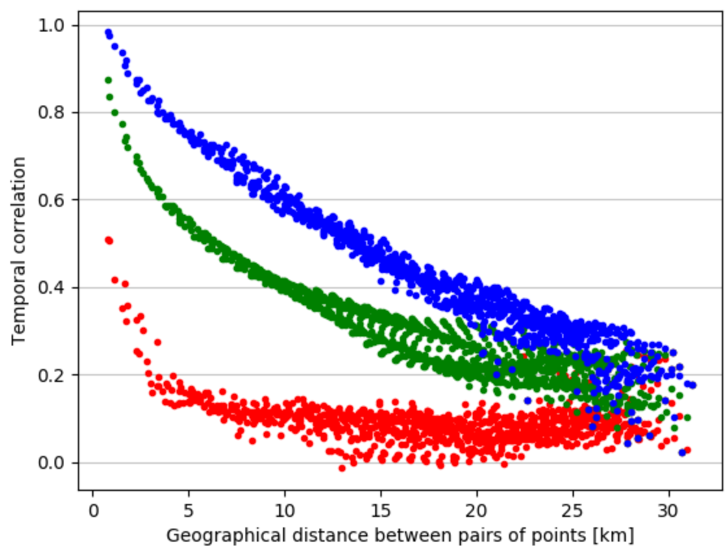

Figure 2. Temporal Correlation versus Geographical distance between each pair of points. Correlation values are grouped into three categories - minimum, medium and maximum values - respectively coloured in red, green and blue.

We can also notice that the minimum correlations for the geographically nearest ten pairs of points are even higher than the maximum correlations for those more distant than 28 km. Such property is an indicator of how well-behaved the relation between temporal correlations and geographical distances in this network structure.

The scatter plot on Figure 3 presents the relation between the euclidean distance and the topological distance between each pair of nodes - the network path with the shortest number of edges between those nodes. We can verify that there is strong linearity in such relation, with a correlation coefficient (R^2) equals to 0.767 and a slope of 1.16. Such a slope value indicates that as the geographical distance increases, the impact is even more significant on the topological distance.

This chart also shows the largest edge in the network (2.5 km), indicated by the maximum geographical distance for the pairs of points within a topological distance of 1 edge. Therefore, there are no pairs of points directly connected in a distance greater than

2.5 km. On the other hand, there are very close nodes, geographically neighbors, but with a high topological distance, up to 12 edges.

The geographical network built up by graph4GIS is introduced in Figure 4. It used a threshold of 0.86, which was the critical threshold for our study case. This output allows us to visualize the structure of network connections spatially.

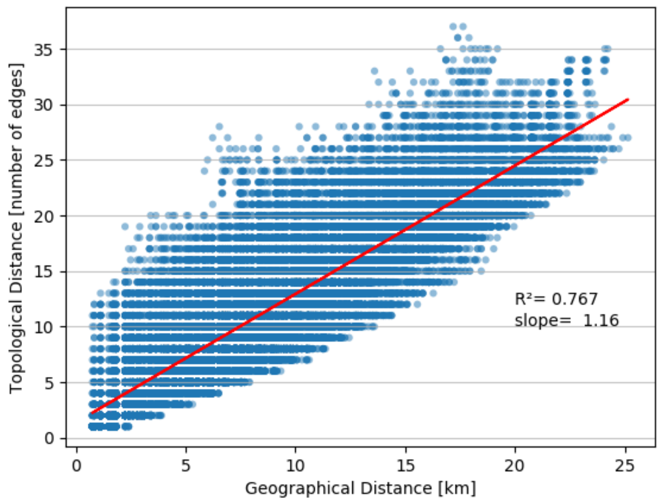

Figure 3. Topological distance versus Geographical (euclidean) distance

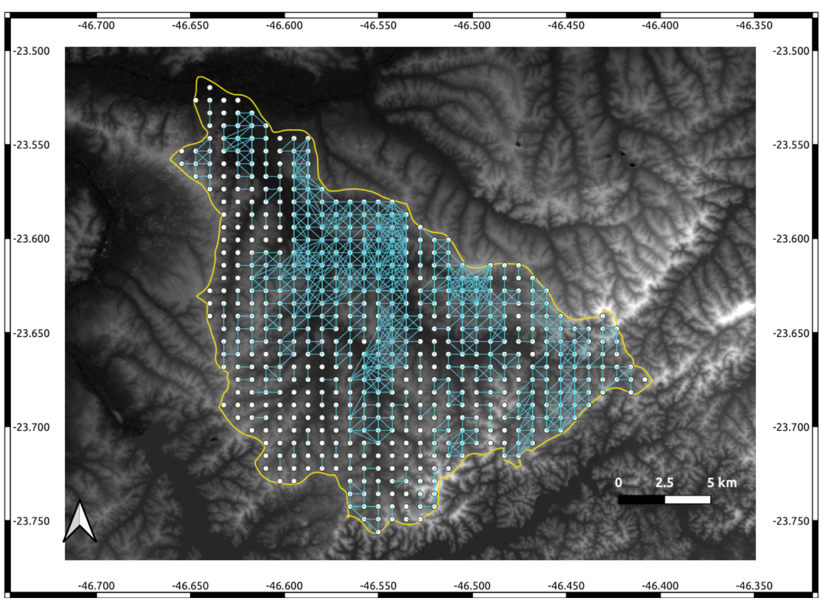

Figure 4. Geographical network for Tamanduateí Basin. The white points represent the nodes, the blue segments are the edges of the network, and the yellow border is the outline of the basin.

4. Final Considerations

This work applied Complex Networks in the study of meteorological networks, aiming to explore topological metrics' behavior in such a context. Based on precipitation time series, this paper introduced some spatial analysis of the system's topological structure.

As a result, we could identify the spatial dependence of temporal correlations, such as the linearity in the relation between the topological and geographical distances between different pairs of points in a hydrological basin. We were also able to verify some peculiarities in the network, such as the maximum geographical length of an edge (2.5 km) and a high maximum topological distance between neighboring nodes (11 edges on the shortest path between nodes closer than 1km to each other).

In future works, we would like to analyze datasets for specific meteorological processes to identify spacial and topological signatures. Besides, we intend to approach larger study areas, including the entire São Paulo Metropolitan Region, and other graph measures, such as degree, clustering coefficient, betweenness, and diameter.

References

[Barabási and Pósfai 2016] Barabási, A.-L. and Pósfai, M. (2016). *Network science*. Cambridge University Press, Cambridge.

[Boers et al. 2019] Boers, N., Goswami, B., Rheinwalt, A., Bookhagen, B., Hoskins, B., and Kurths, J. (2019). Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature*, 566(7744):373–377.

[Ceron et al. 2019] Ceron, W., Santos, L. B. L., Dolif Neto, G., Quiles, M. G., and Candido, O. A. (2019). Community Detection in Very High-Resolution Meteorological Networks. *IEEE Geoscience and Remote Sensing Letters*, pages 1–4.

[DECEA 2010] DECEA (2010). Ministério da defesa comando da aeronáutica estado-maior da aeronáutica. *Ministério da Defesa Comando da Aeronáutica*, pages 1–24.

[Ramalho 2007] Ramalho, D. (2007). Rio Tamanduateí - nascente à foz: percepções da paisagem e processos participativos. *Paisagem e Ambiente*, (24):99.

[Redemet 2015] Redemet (2015). Redemet. `https://www.redemet.aer.mil.br`.

[Santos et al. 2019] Santos, L. B. L., Carvalho, L. M., Seron, W., Coelho, F. C., Macau, E. E. N., Quiles, M. G., and Monteiro, A. M. V. (2019). How do urban mobility (geo)graph's topological properties fill a map? *Appl Netw Sci*, 4(91).

[Santos et al. 2017] Santos, L. B. L., Carvalho, T., Anderson, L. O., Rudorff, C. M., Marchezini, V., Londe, L. R., and Saito, S. M. (2017). An rs-gis-based comprehensiveimpact assessment of floods—a case study in madeira river, western brazilian amazon. *IEEE Geoscience and Remote Sensing Letters*, 14(9):1614–1617.

[Santos et al. 2017] Santos, L. B. L., Jorge, A. A. S., Rossato, M., Santos, J. D., Candido, O. A., Seron, W., and de Santana, C. N. (2017). (geo)graphs - Complex Networks as a shapefile of nodes and a shapefile of edges for different applications. 321124491(November).

[Tsonis et al. 2006] Tsonis, A. A., Swanson, K. L., and Roebber, P. J. (2006). What do networks have to do with climate? *Bulletin of the American Meteorological Society*, 87(5):585–595.