An Approach to Visualization and Clustering-based Analysis on Spatiotemporal Data

Maxwell Guimarães de Oliveira, Cláudio de Souza Baptista

Departamento de Sistemas e Computação, Universidade Federal de Campina Grande (UFCG) - Brazil
maxwell@ufcg.edu.br, baptista@dsc.ufcg.edu.br

Abstract. Currently, there is a considerable amount of spatiotemporal data available in various media, especially on the Internet. The visualization of spatiotemporal data is a complex task that requires suitable visual resources that can enable users to have a correct interpretation of the data. Apart from the use of visualization techniques, the use of techniques of knowledge discovery in databases has proven to be relevant for the exploratory analysis of spatiotemporal data. The state-of-the-art in the visualization of spatiotemporal data leads to the conclusion that the area is still deficient in solutions for the viewing and analysis of those data. Many approaches cover only spatial issues, ignoring the temporal characteristics of such data. In this context, the main objective of this research work is to improve the user experience in spatiotemporal visualization and analysis, going beyond the universe of the visualization of spatiotemporal raw data by considering the importance of the visualization of spatiotemporal data derived from a knowledge discovery process, more specifically, clustering algorithms. This goal is achieved by defining an innovative approach for the analysis and visualization of spatiotemporal data, and its implementation, called GeoSTAT (Geographic Spatiotemporal Analysis Tool). GeoSTAT includes important features of the main existing approaches and adds specific visualization techniques that are geared to the temporal dimension and the use of clustering algorithms, enhancing unexplored features in the spatiotemporal data. The validation of the proposed approach occurs through a case study that addresses spatiotemporal data from a specific domain, to demonstrate the end-user experience of the visualization techniques that are combined in the proposed approach presented in this article.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications—*Spatial databases*; *GIS*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Clustering*

Keywords: clustering, data analysis, spatiotemporal data, visualization

1. INTRODUCTION

The widespread use of devices that capture a geographic location, such as smartphones and GPS devices installed in automobiles, has generated large amounts of information that concern time and space, such as the trajectory of mobile objects, fire hot spots, occurrence of dengue outbreaks, atmospheric discharges, and criminality maps. This considerable volume of spatiotemporal data is available in a variety of media types, specially on the Internet. In so much information, it is necessary to provide decision support systems and analytics, which can help decision-making users to extract relevant knowledge, intuitively and quickly, such as information that is relevant to the prediction of future events.

Visualization techniques are widely known as being powerful in the decision-making process [Johnston 2001] because they take advantage of human capabilities to rapidly notice and interpret visual patterns [Andrienko et al. 2003; Kopanakis and Theodoulidis 2003]. However, we know that the spatial visualization resources supplied by most of the existing geographic information systems are not sufficient for decision support systems [Bédard et al. 2001].

The visualization of spatiotemporal data is a complex task that requires the use of appropriate visual resources that allow users to have a correct interpretation of the information under analysis. The visualization and analysis of spatiotemporal data are tasks that have gained prominence in several areas, such as biology, electrical power transmission, urban tracking, criminology, and civil construction.

Moreover, spatiotemporal data present strong challenges to data analysts. First, there is the complexity of the spatial dimension, which requires human capabilities to determine the spatial relationships [Andrienko et al. 2008]. Second, there is the modeling of the temporal dimension. According to Andrienko et al. [2010], it is necessary to address time in an efficient manner when performing spatiotemporal visualization. The understanding that space and time are inseparable and that there is nothing spatial that is not temporal must permeate the research on spatiotemporal visualization. A reasonable solution in visualization and the analysis of spatiotemporal data should offer at least the following: resources for treating both the spatial and temporal dimensions (spatiality and temporality); domain independence (generality); freedom for the user to handle the visualized data and apply filters (flexibility); a connection with several data sources in a practical and efficient manner (interoperability); and data mining based on spatiotemporal clustering (mining).

It is essential to provide to the users resources for handling both the spatial and the temporal dimensions in a spatiotemporal data analysis system. The singularities in any of these dimensions must not be discarded because they could reveal implicit relationships that match the reality of the analyzed data. Furthermore, the use of spatiotemporal data mining algorithms integrated with modern data visualization techniques improves the usability for the decision maker when analyzing large spatiotemporal datasets.

Nonetheless, most existing spatiotemporal visualization systems do not address appropriately the temporal dimension because they focus on spatial visualization. Therefore, an important research issue is how to offer temporal operators that, used with the spatial data operators, can improve the experience of the end users, who are interested in performing visual analysis on spatiotemporal data.

Within this context, this article presents a new approach to visualize and analyze spatiotemporal data. This approach accounts for the six essential characteristics proposed by Andrienko et al. [2010], as mentioned previously. We then present a new system based on this proposed approach, called GeoSTAT (Geographic SpatioTemporal Analysis Tool), for the visualization and analysis of spatiotemporal data.

A case study using the GeoSTAT system and, of course, applying the proposed approach, was proposed to perform a spatiotemporal analysis using data on fire hot spots and failure events in power transmission lines and is aimed at finding evidence that supports the hypothesis that fires that occur close to transmission lines could be the cause of failure events in the power system.

The remainder of this article is organized as follows. Section 2 focuses on the presentation of the proposed approach and the proposed system. Section 3 addresses a case study to validate the proposed ideas. Section 4 discusses related work. Finally, section 5 concludes the article and presents further work to be undertaken.

2. VISUALIZATION AND CLUSTERING-BASED ANALYSIS ON SPATIOTEMPORAL DATA

This section presents our approach for spatiotemporal visualization. This approach joins spatial and temporal visual analysis and clustering techniques on spatiotemporal data.

2.1 Visualization Techniques

Visualization is a mechanism by which human beings interact with computers to facilitate data analysis. In this subsection, we address spatial, temporal and spatiotemporal data visualization techniques that are used to facilitate end user interaction.

Concerning the visualization in the spatial dimension, our approach provides the following resources: Dynamic Map Viewer, Map Layers and Spatial Query. The Dynamic Map Viewer enables us to visualize spatiotemporal data, together with the spatial relationships: the neighborhood, distance, and intersection.

Map Layers enables us to overlay layers for the purpose of improving the data analysis. Using a spatial query, users can pose queries that use both spatial and conventional attributes, through topological operators such as inside, meets, covers, crosses, and overlaps.

The temporal dimension includes the following capabilities: Temporal Slider, Temporal Distribution Chart, and Temporal Query. By using the temporal controller, users can determine either a timestamp or a temporal interval to visualize the data. Temporal animation is also provided to enable the visualization of dynamic maps through different timestamps, to allow spatiotemporal changes to be visualized. The Temporal Distribution Chart is another important visualization tool in which users can observe the periodicity (weekly, monthly, annually) of the data. This chart presents timestamps in which a given event occurs. The Temporal Query enables us to process data based on temporal relationships such as Begin, End, and Between.

In our approach, we provide cluster-based algorithms for spatiotemporal data to enhance the user's experience in spatiotemporal data visualization. Clustering algorithms can reveal distribution patterns on spatiotemporal data. Hence, users can perform a spatiotemporal neighborhood analysis.

It is also possible to find regions and periods that have great density, in other words, the portions of the data that concentrate on the spatial and temporal dimensions. Furthermore, events that rarely occur in given regions can also be detected.

We do not force the use of a specific clustering algorithm. Instead, we consider the advantages of using data mining techniques on visualizing spatiotemporal data, regardless of which clustering algorithm is chosen. It is possible to perform spatiotemporal clustering on two different layers, aiming to visualize two distinct events that are close in space and time. This correlation could reveal a cause-and-effect relation on the underlying data, helping users to minimize the occurrence of an event in the function of another one.

2.2 The Geographic Spatiotemporal Analysis Tool - GeoSTAT

This subsection introduces GeoSTAT (Geographic Spatiotemporal Analysis Tool), which is a new web-based system for spatiotemporal visualization and analysis and is an implementation of our proposed approach on the visualization and analysis of the spatiotemporal data.

Through the GeoSTAT system, the user interested in viewing and analyzing a spatiotemporal dataset will be able to use several visualization resources that address both spatial and temporal dimensions. Moreover, clustering-based data mining algorithms, adapted for the spatiotemporal domain, were integrated into the system. In addition to the advantages of being a web application, GeoSTAT was conceived with a general point of view. For this reason, it is a domain-independent system, which can be connected to any spatiotemporal data source available over the Web by implementing the spatial data sharing services standardized by the Open Geospatial Consortium[1] (OGC).

2.2.1 *GeoSTAT Components.*

The interactive user interface of the GeoSTAT system is comprised of ten components that are responsible for the functionalities offered by the system. Figure 1 presents this interface and enumerates

[1]OGC - Open Geospatial Consortium. More information in: http://www.opengeospatial.org/

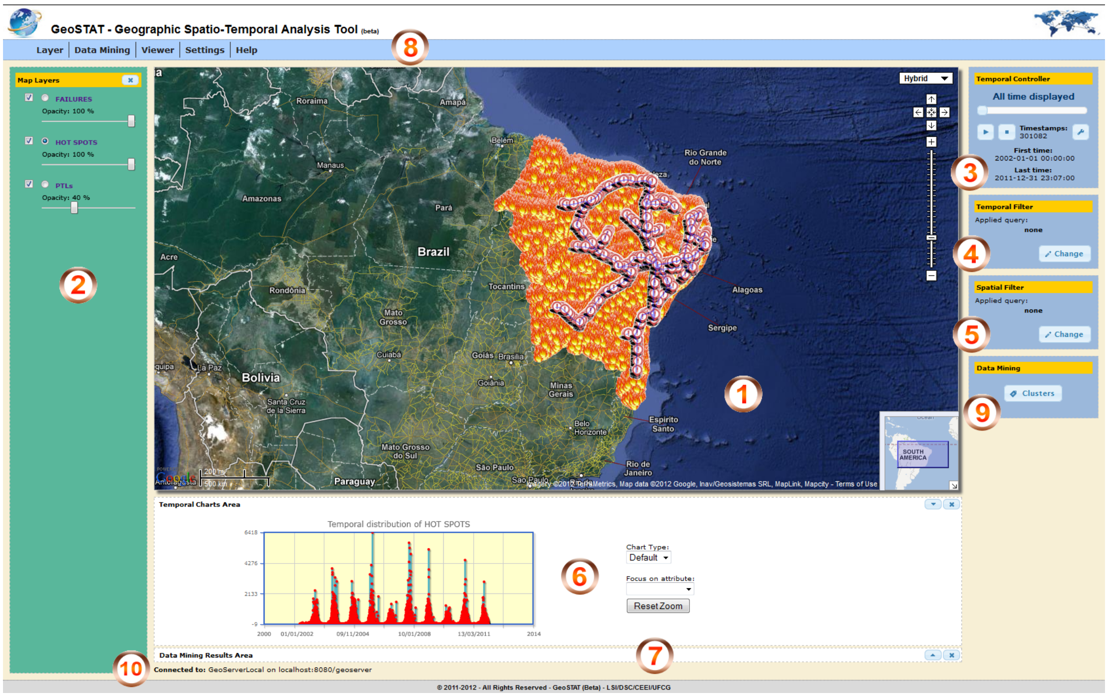

Fig. 1. The main interface and components of GeoSTAT system displaying data layers used in case study presented in section 4.

these components: 1) map; 2) spatiotemporal layers (overlap); 3) temporal controller; 4) temporal filter; 5) spatial filter; 6) temporal distribution graphic; 7) data mining results; 8) actions menu; 9) data mining; and 10) information about the connected data servers.

The map component uses the Google Maps API to offer a dynamic map. The spatiotemporal layers component allows users to add layers and spatiotemporal (or only spatial) data published in servers that implement the OGC WMS (Web Map Service) and WFS (Web Feature Service) services. These data are plotted on the map and are made available through the components that address the temporal dimension, such as the temporal controller, the temporal filter and the temporal distribution graphic. They are also made available for clustering-based data mining through the system.

Through the use of the temporal controller, it is possible to change the map visualization using a temporal filter. This filter can be defined as either a given instant (timestamp) or a more abstract level of temporal resolution, such as months, for example. The temporal controller also allows the production of a temporal animation, which allows the user to visualize on the map the eventual changes in the spatial distribution of the data as a function of the temporal variation. It also displays a specific timestamp and enables the observation on the map of a spatial distribution of data at this specific timestamp. At any time, it can terminate the animation and view the spatial distribution of the whole dataset on the map again, regardless of the temporal dimension.

In addition to the temporal controller, another available temporal visualization resource is the temporal distribution graphic which is responsible for helping the user to visualize changes in the spatiotemporal data as a function of time, adding to the map resource, which helps the visualization of the distribution as a function of space.

The spatial and temporal filter components are responsible for the spatial and temporal query and selection, respectively, of the data visualized through the spatiotemporal layers. Through the temporal filter, the user can, by means of four filter options and observing the temporal resolution used, reduce the spatiotemporal dataset for visualization and analysis. The four options available for the temporal filter are: from, until and between. On the other hand, through the spatial filter, it is possible to visualize a topological relationship between two spatial or spatiotemporal layers previously added to the system, regardless of the source data source. It is possible to perform the following topological relations between two layers: intersects, contains, crosses, touches, covers and overlaps. It is also possible to apply negation (not) to each one of these relations, in cases in which negations are relevant for the analysis performed by the user.

In the component of data mining, it is possible to perform clustering-based data mining in the previously added layers, to view the result of a previous data mining process and the detailed status of data mining processes under execution. The data mining processes run in the background; as a result, users do not need to wait for the end of this processing because they could perform other tasks.

The component of data mining results is responsible for offering the statements necessary for the spatiotemporal visualization and for browsing a layer that contains the data mining results. The user could browse through the timestamps that have the occurrences of clusters and view each cluster separately on the map. If the data mining process is performed using two layers, the user will have the option of viewing only the relevant clusters, in other words, those clusters that have at least one point in each layer as well as options to view only the clusters that group points that are only in one layer. It is also possible to see all of the clusters that are in a given timestamp or even all of the clusters in total.

Finally, the actions menu component offers shortcuts for the remaining components of the interactive graphic interface of the GeoSTAT system, and the connected source data server component is responsible for displaying information about the data servers that are connected to a user's session of the system.

2.2.2 *GeoSTAT Architecture.*

The GeoSTAT system architecture is defined using three layers: visualization, control and persistence. The visualization layer is responsible for the user interface, providing components for loading, handling and visualizing the data through the temporal and spatial dimensions, presented in subsection 2.2.1.

The control layer is responsible for the processing of all of the requests that are generated and sent from the visualization layer, besides being responsible for the communication with the persistence layer; therefore, the control layer is the kernel of the GeoSTAT system. Figure 2 presents the five existing modules in the control layer. These modules are activated according to the nature of the request to be processed by this layer.

The request interpretation module (Figure 2) is the main module of the control layer. It is responsible for receiving and treating every request that comes from the visualization layer and for establishing contact with the other modules, besides making contact with the persistence layer. There are two types of treatment to the requests that arrive at the request interpretation module: query or data delivery requests and data processing requests. The data requests are sent directly to the persistence layer, which is responsible for interpreting and processing this type of request. On the other hand, the data processing requests can be forwarded to the data mining module or to the spatial query module.

The spatial query module (Figure 2) is responsible for the processing of spatial queries between two different layers. The result of the query processing (spatial filter) is sent to the visualization layer, to be exhibited to the end user.

The data mining module integrates several known clustering algorithms. These algorithms were obtained from the Weka toolkit [Hall et al. 2009]. Seven algorithms were adapted and are available on the GeoSTAT system: COBWEB, DBScan, K-Means, X-Means, Expectation-Maximization,

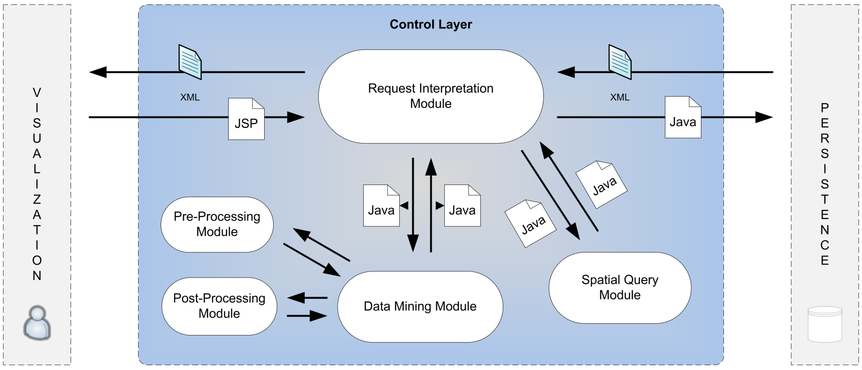

Fig. 2. Control modules of the GeoSTAT system architecture.

Farthest-First and OPTICS. Hence, the GeoSTAT system is capable of performing clustering-based spatiotemporal data mining on any spatial or spatiotemporal database. The output returned by the data mining module is stored in a spatiotemporal database and made available for query from the system as soon as the processing is complete. The data mining module uses threads for concurrent processing.

To make possible the spatiotemporal integration and adaption of the several data mining algorithms used, we developed data pre-processing and post-processing modules. These auxiliary modules are responsible for preparing data to be used by the algorithm selected by the user and for preparing the results obtained through the execution of this algorithm for treatment by the visualization layer. Figure 3 details the data mining module.

The pre-processing module is responsible for reading the dataset from a WFS GetFeature request using the GML format. Then, the dataset is transformed into a temporary file using the CSV (Comma Separated Values) format. The CSV file is used by the data mining clustering algorithm, which is selected as an input parameter.

The post-processing module starts after the data mining process is concluded. This module prepares the results from the data mining process to be analyzed in the GeoSTAT visualization layer.

The data mining module controls the choice of the clustering algorithm and separates the input data into groups for data mining in the temporal dimension. It manages the execution of 12 concurrent

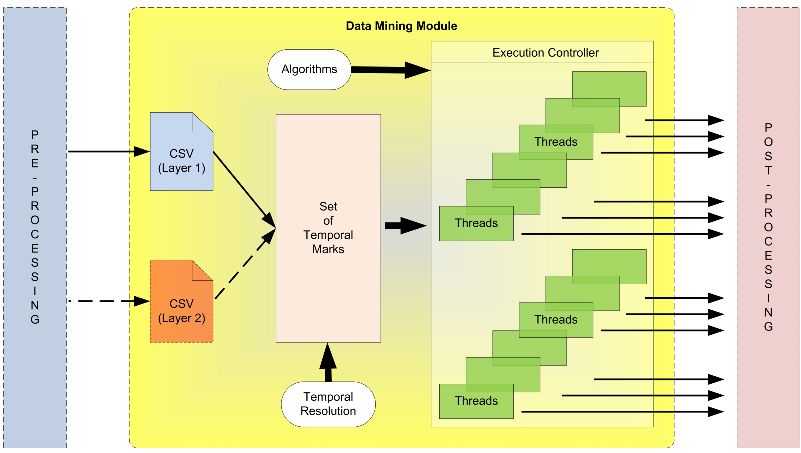

Fig. 3. Working Schema for Data Mining Module.

threads. Each thread executes the clustering algorithm with a group of input data, to optimize the cost of the data mining process as a whole. At the end of each thread, the result clusters are forwarded directly to the post-processing module. Hence, this process continues until there is no data to process.

The persistence layer is responsible for connecting the GeoSTAT system to the databases requested by the users through the components of the visualization layer. When a data request is received from the control layer, the persistence layer first identifies the type of connection that will be established. The GeoSTAT can connect either to the OGC WMS and WFS services or to a spatiotemporal database developed to operate exclusively with the system. The OGC services are accessed from their web servers.

The spatiotemporal database stores information used by the GeoSTAT system to connect to the OGC services as well as the complete results of the data mining processes performed by the system and available for visualization.

3. CASE STUDY

This study is composed of the analysis of two sets of spatiotemporal data. Each set is comprised of records of a spatiotemporal event.

3.1 Data

To conduct this study, we used georeferenced spatiotemporal data about fires detected in the Northeastern region of Brazil, which were supplied by the National Institute for Space Research[2] (INPE) through the Weather Forecast and Climatic Studies Center (CPTEC), which publishes this type of information daily through their Fire Monitoring Portal[3].

We obtained a total of 2,361,040 records of fire hot spots detected in the region, in the period between 01-01-2002 and 12-31-2012. The spatiotemporal data were obtained in the ESRI Shapefile format using the WGS84 geographic reference system, and the temporal data were obtained according to the GMT. According to INPE, their system detects the presence of fire in the vegetation, and the mean error in the spatial location of the spots is approximately 400 meters, with a standard deviation of approximately 3 kilometers and with approximately 80% of the spots detected in a distance of one kilometer from the coordinates indicated by the system. In terms of temporal validity, the satellites offer a mean temporal resolution of three hours. This resolution is the mean time between the passes of two satellites that capture information about the same region.

Another spatiotemporal database was used in this study. This database concerns failure events in power transmission lines and is recorded by the San Francisco Hydroelectric Company (Eletrobrás/Chesf), which operates throughout the Northeastern region of Brazil. Since we could not obtain official data from Eletrobrás/Chesf, due to technical and confidentiality matters, we developed an algorithm to generate spatiotemporal failure events randomly, under spatial constraints imposed by Eletrobrás/Chesf's transmission line network and the temporal constraints imposed by the other database used in this study.

We generated a total of 131,834 failure records in Eletrobrás/Chesf's transmission lines, in the period between 01-01-2002 and 12-31-2012. These records were stored in a spatiotemporal database, which was also in the WGS84 geographic reference system, with temporal information according to the GMT. Aiming to help in the visual analysis of the transmission line failure events, we also used a set of spatial data that contained Eletrobrás/Chesf's transmission line network.

[2]INPE - Brazilian National Institute for Space Research. More information in: http://www.inpe.br/
[3]INPE/CPTEC - Fires Monitoring Portal. Available in: http://www.inpe.br/queimadas/

Both datasets used in this study share the same spatial geometry (POINT) and also the same temporal resolution (timestamp). To use the data in the GeoSTAT system, we installed the Geoserver web map server and created layers for each dataset.

To conduct this study, the GeoSTAT system user will be called an analyst, a specialist user in the domain who is looking for relevant information that is contained implicitly in a large volume of spatiotemporal data.

3.2 Experiment

Figure 1 shows the GeoSTAT system interface with the three spatiotemporal layers loaded into the system from the data connection with Geoserver. What is seen is the result of approximately two million and a half points plotted in the map, which is enough to fill the whole Northeastern region.

The temporal distribution graphics, generated and shown automatically when a spatiotemporal layer is loaded and selected in the GeoSTAT system, allows the analyst to verify the behavior of the whole volume of data. By observing the graphic that corresponds to the fire hot spots layer (shown in Figure 1), we note that there is an annual repetition of the distribution of the number of spots detected, where the maximums concentrate in the first and the last months of each year. This time interval is the period when the Northeastern region registers the highest temperatures, which contributes to the occurrence of new fire hot spots. Through this graphic, we can also observe that the maximum number of spots detected in one day, during the 10-year period, was 6,418 spots. This number was reached on 11-07-2005.

By observing the graphic that corresponds to the transmission line failures layer, we notice a temporal behavior that is almost continuous. Once the data were randomly generated through an algorithm, the temporal distribution of the occurrences was uniform, registering a maximum of three occurrences in one single day.

For a better visualization of the power line failures and of the detected fire hot spots, the systemcould use a more generic temporal resolution than the timestamp, such as "Date and Time", for example, which would allow the user to join all of the records that occur between "10-15-2011 15:00:00" and "10-15-2011 15:59:59" into one single view, for example. However, the cost would be too high for the analyst to view image by image, time by time, manually, to find interesting behaviors. The use of the clustering technique emerges as a good option to reduce the cost to the analyst, by making the spatiotemporal clustering of the events.

With the layers "FAILURES" and "HOT SPOTS" added to the GeoSTAT system, we activate the spatiotemporal clustering option offered by the system to perform data mining on both layers. This option enables the analyst to view the spatiotemporal clusters of each separate event and the relevant clusters, in other words, the spatiotemporal clusters that contain records of both events.

To execute the data mining, aside from the three input layers, the user had to inform the required parameters: "Date + 3-3 hours" for temporal resolution, and DBScan was used as the data mining algorithm, with MinPoints = 2 and Epsilon = 0.013472.

The choice of the value 0.013472 for the Epsilon parameter of DBScan is because one second (an angular measurement unit) is approximately equal to 30.9 meters. Because approximately 80% of the fire hot spots detected by INPE occur within one kilometer from the indicated coordinates, and the mean error in the spatial location of the records is 400 meters, we thought it was reasonable that the radius of a generated cluster ranged from 1 to 1.5 kilometers. Because 48.5 seconds is approximately equal to 1,498.65 meters (1.5 kilometers) and one decimal degree has 60 minutes and 60 seconds, we concluded that 1,498.65 meters is approximately equal to 0.013472 meters.

3.3 Results and Conclusions

The data mining process in this case study spent seven hours, 37 minutes, and five seconds. It was executed in a web application server, running the Microsoft Windows 7 Professional (64-bit) operating system, with an Intel Core i7 processor and 16 GB of RAM.

The statistical results for the classification of the records after the execution of the algorithm showed that approximately 86% of the records were associated with a spatiotemporal cluster. The remaining records, approximately 14% of the total, were considered to be outliers because they did not belong to any spatiotemporal cluster and represented only isolated occurrences in space-time. Among the records associated with a spatiotemporal cluster, only 1.29% of the whole dataset (32,275 records) were considered to be relevant by the GeoSTAT system. These results mean that only these records are contained in relevant spatiotemporal clusters, those which contain records of both of the studied events.

From the 318,901 spatiotemporal clusters generated, only 1,376 (0.43%) were considered to be relevant under the viewpoint of the measurement parameters used in the execution of the data mining algorithm. Each irrelevant cluster grouped, on average, 6,623 records, while each relevant cluster grouped, on average, only 23 records.

Figure 4 presents a screenshot captured from the GeoSTAT system, which shows in the map all of the relevant spatiotemporal clusters generated for the 10-year period of the dataset. In this visualization, the analyst can notice that most of clusters are concentrated at Southeast of the state of Ceará, more precisely at the border of the states Ceará, Paraíba and Rio Grande do Norte; this region is highlighted in the picture.

The generated spatiotemporal clusters can be browsed with the components for temporal selection and, from this definition, with the individual selection of each cluster corresponding to the previously selected timestamp. The analyst can choose the visualization of relevant clusters only or the visualization of all of the clusters. The analyst can also visualize each individual cluster or visualize all of the clusters, regardless of the temporal dimension.

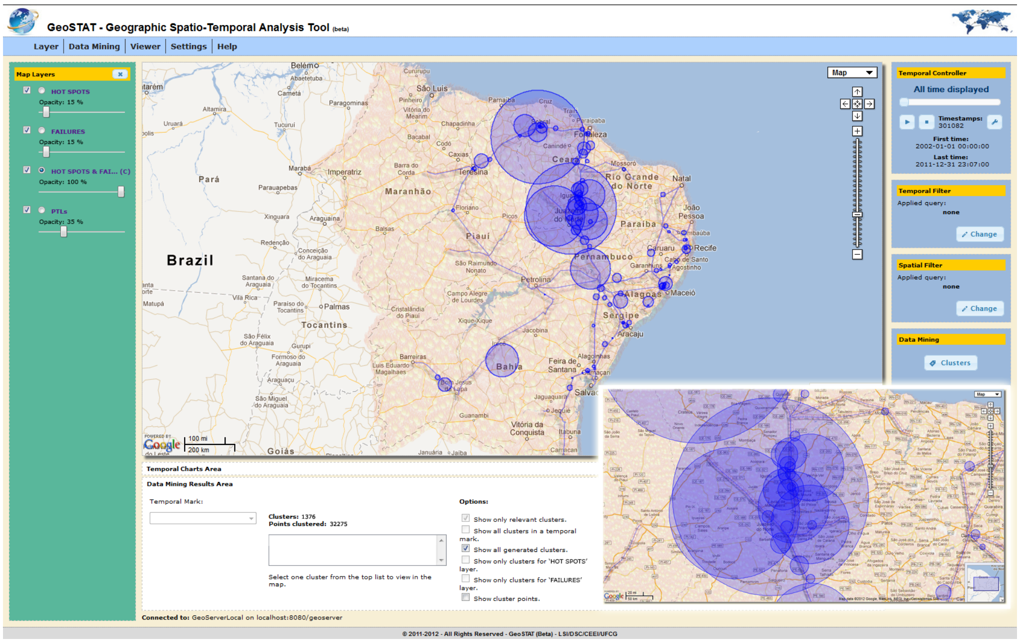

Fig. 4. GeoSTAT system showing all the relevant clusters.

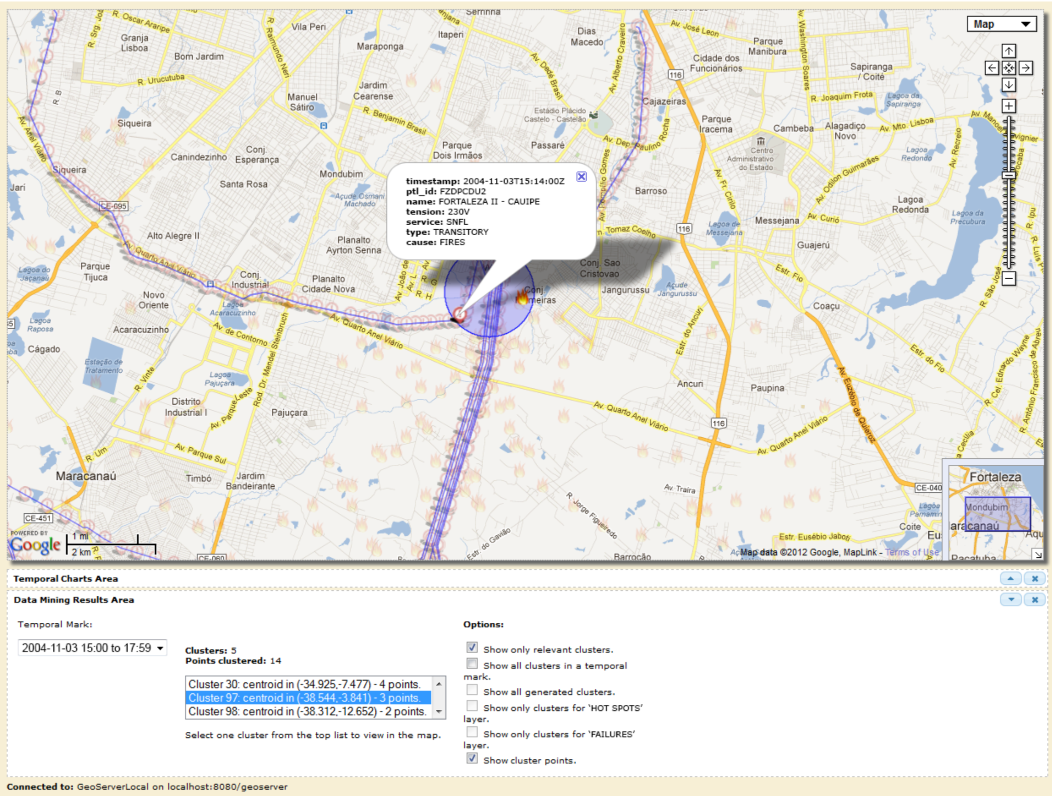

Fig. 5. GeoSTAT system displaying, in detail, the spatiotemporal cluster no. 97, with timestamp "11-03-2004 03:00 p.m. to 05:59 p.m.".

For the analyst interested in confirming the hypothesis that some of the fire hot spots are the cause of failures in power transmission lines, Figure 5 exemplifies a case in which the hypothesis is confirmed. A failure that occurs in the line "FORTALEZA II - CAUIPE" at 03:14 p.m. in 11-03-2004 had its cause specified as "FIRES" and, moreover, because of the data mining performed together with data from records of fire hot spots detected in that region at the same period as the failure, noted a spatiotemporal clustering between this failure and two fire hot spots: one detected at 04:08 p.m., with an approximate distance of 1 kilometer from the failure, in the East direction, and another one, detected at 04:01 p.m., with an approximate distance of 1.5 kilometers from the failure, in the North direction. If we consider the spatial precision errors and the temporal resolution of these data, the analyst could point to these two fire hot spots as the actual causes of the failure.

The results achieved with the use of the GeoSTAT system were satisfactory for the applications domain explored in this study. The visualization resources explored allowed the discovery of interesting implicit information, from two large volumes of data.

It is important to observe that the statistical data mining results pointed to an index of relevant clusters that most specialists on this type of event would expect. These results are mainly due to the use of simulated records of power transmission line failures. The use of real data, captured and structured by Eletrobrás/Chesf, will certainly produce better results, such as the presence of more relevant clusters.

In addition to using real data, the specialists have made available, through the GeoSTAT system, several spatiotemporal clustering algorithms. Their results can be compared and analyzed to find new relevant information.

4. RELATED WORK

This section addresses related studies that concern the visualization and analysis of spatiotemporal data.

Ferreira et al. [2011] propose an interactive visualization system that supports the visual analysis of spatiotemporal bird distribution models. This system is a spatiotemporal approach that involves the specific domain of birds. It is important to highlight that besides being valid for only one specific domain, the solution does not provide mechanisms to connect to external databases and is currently constrained to the database developed by the authors.

Roth et al. [2010] present a web-mapping application that supports spatiotemporal exploration in the criminology domain. This application offers a spatiotemporal browsing resource that animates simultaneously a map and a frequency histogram that illustrates the temporal distribution. This application enables the visualization of the variation of data through time, organized into crime categories. Although this solution supports spatiotemporal data, it is limited to one specific application domain, and there is no database interoperability.

Compieta et al. [2007] propose a domain-independent spatiotemporal data-mining system that is based on association rules. They propose two visual tools: one based on Google Earth, which is georeferenced, and the other based on Java-3D, which is not georeferenced. Nonetheless, the visual support for the temporal dimension is too limited, as the proposed tools enforce the spatial dimension.

Reda et al. [2009] developed a visual exploration tool to analyze changes in groups of dynamic spatiotemporal social networks. They propose two interesting techniques for spatiotemporal visualization. The affiliation timeline displays the structure of the community in the population and its evolution over time, and the spatiotemporal cube enables the visualization of the movement of communities in a spatial environment. However, besides being valid only for the domain of social groups, it does not describe how the user should supply the data for visualization and analysis. We conclude that this solution has some limitations that concern data heterogeneity.

Andrienko et al. [2007] address a framework for the visual analysis of spatiotemporal data that represents the trajectory of mobile objects. This framework combines database operations with computational processing, data mining and interactive visual interfaces. The solution highlights the use of the OPTICS clustering algorithm for the detection of frequently visited places and database reduction. It is a domain-independent solution, although it is constrained to the trajectory of mobile objects that are represented by points in space. Moreover, the authors do not make clear the acceptable format for the trajectory data.

Chen et al. [2003] presents a visualization tool for spatiotemporal data that integrates GIS and Self Organizing Map technique (SOM) [Kohonen 2001], for spatiotemporal clustering. However, this tool is designed for the domain of crimes. Hence, no interdomain interoperability is provided.

Among the previously mentioned research studies, which focus on the visualization and analysis of spatiotemporal data, some address domain-specific solutions and, thus, are useful for a limited group of users. Furthermore, many of them do not provide flexibility concerning the use of heterogeneous datasets, often requiring a considerable effort from users to adapt their datasets to the chosen application in order to perform the analysis. Additionally, there are problems that concern usability because the user interfaces do not provide enough freedom to the end users to include or remove feature types that they might find to be relevant to their tasks.

5. CONCLUSIONS AND FUTURE WORK

In this article, we presented the proposal of a new approach and a new system for the visualization and analysis of spatiotemporal data. This approach addressed the six features needed by a solution for spatiotemporal visualization and analysis: resources for the spatial dimension, resources for the temporal dimension, domain independence, flexibility, interoperability, and data mining based on spatiotemporal clustering. This solution prioritizes the end user, offering a set of functionalities that allow the execution of a job in a practical and efficient manner.

Finally, we conclude that the proposed approach met its objectives, proving to be satisfactory and efficient. We also conclude that many improvement issues for the proposed system can be addressed in future studies, which certainly will contribute to a more robust system. One such improvement could be the inclusion of another data mining technique, such as spatiotemporal association rules.

REFERENCES

Andrienko, G., Andrienko, N., Demsarb, U., Dranschc, D., Dykes, J., Fabrikant, S. I., Jernf, M., Kraakg, M.-H., Schumannh, H., and Tominskih, C. Space, Time and Visual Analytics. *International Journal of Geographical Information Science* 24 (10): 1577–1600, 2010.

Andrienko, G., Andrienko, N., Dykes, J., Frabrikant, S. I., and Wachowicz, M. Geovisualization of Dynamics, Movement and Change: key issues and developing approaches in visualization research. *Information Visualization* 7 (3-4): 173–180, 2008.

Andrienko, G., Andrienko, N., and Wrobel, S. Visual Analytics Tools for Analysis of Movement Data. *ACM SIGKDD Explorations Newsletter* 9 (2): 38–46, 2007.

Andrienko, N., Andrienko, G., and Gatalsky, P. Exploratory Spatio-temporal Visualization: an analytical review. *Journal of Visual Languages & Computing* 14 (6): 503–541, 2003.

Bédard, Y., Merrett, T., and Han, J. Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery. In H. J. Miller and J. Han (Eds.), *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, New York, NY, USA, pp. 53–73, 2001.

Chen, H., Atabakhsh, H., Petersen, T., Schroeder, J., Buetow, T., Chaboya, L., O'Toole, C., Chau, M., Cushna, T., Casey, D., and Huang, Z. COPLINK: visualization for crime analysis. In *Proceedings of the Annual National Conference on Digital Government Research*. Boston, MA, USA, pp. 1–6, 2003.

Compieta, P., Martino, S. D., Bertolotto, M., Ferrucci, F., and Kechadi, T. Exploratory Spatiotemporal Data Mining and Visualization. *Journal of Visual Languages & Computing* 18 (3): 255–279, 2007.

Ferreira, N., Lins, L., Fink, D., Kelling, S., Wood, C., Freire, J., and Silva, C. BirdVis: visualizing and understanding bird populations. *IEEE Transactions on Visualization and Computer Graphics* 17 (12): 2374–2383, 2011.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. The WEKA Data Mining Software: an update. *ACM SIGKDD Explorations Newsletter* 11 (1): 10–18, 2009.

Johnston, W. L. Model Visualization. In U. Fayyad, G. G. Grinstein, and A. Wierse (Eds.), *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann Publishers, San Francisco, CA, USA, pp. 223–227, 2001.

Kohonen, T. *Self-Organizing Maps*. Information Sciences. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.

Kopanakis, I. and Theodoulidis, B. Visual Data Mining Modeling Techniques for the Visualization of Mining Outcomes. *Journal of Visual Languages and Computing* 14 (6): 543–589, 2003.

Reda, K., Tantipathananandh, C., Berger-Wolf, T., Leigh, J., and Johnson, A. E. SocioScape - a Tool for Interactive Exploration of Spatiotemporal Group Dynamics in Social Networks. In *Proceedings of the IEEE Information Visualization Conference*. Atlantic City, NJ, USA, pp. 1–2, 2009.

Roth, R. E., Ross, K. S., Finch, B. G., Luo, W., and MacEachren, A. M. A User-Centered Approach for Designing and Developing Spatiotemporal Crime Analysis Tools. In *Proceedings of GIScience*. Zurich, Switzerland, pp. 66–71, 2010.