

## A Web-based Environment for Analysis and Visualization of Spatio-temporal Data provided by OGC Services

Maxwell Guimarães de Oliveira, Cláudio de Souza Baptista and Ana Gabrielle Ramos Falcão

Laboratory of Information Systems - Computer Science Department

Federal University of Campina Grande (UFCG)

Campina Grande, Brazil

E-mails: maxwell@ufcg.edu.br, baptista@dsc.ufcg.edu.br, anagabrielle@copin.ufcg.edu.br

**Abstract**—The popularity of GPS devices has led to the quick increasing of spatial data volume in the web. Although there are several studies on spatial data sets, not many deal with the temporal variation that may exist on these sets. Most of the approaches that implement a visual analysis of spatiotemporal data still reveal limitations regarding its flexibility, usability and, mostly, generalization. In order to improve these limitations, we propose a new approach, a web-based spatiotemporal viewer and analyzer, which is domain-independent, deals with the temporal variation, and may be connected to any map server that implements the Web Map Service and Web Feature Service, specified by the Open Geospatial Consortium, and thus, it promotes interoperability. Furthermore, our approach includes data mining clustering algorithms, providing an intuitive visual analysis of the results. We performed a case study to validate the proposed solution and the improvements in spatiotemporal visual analysis. The results showed that the new proposed approach facilitates the analysis of spatiotemporal data by humans.

**Keywords**—visualization; analysis; spatiotemporal; data mining; OGC services.

### I. INTRODUCTION

The constant growth of the use of GPS-based (Global Positioning System) devices, such as smartphones, along with the ease of sharing the information from such devices in the internet have substantially increased the volume of spatiotemporal data in the web.

Such spatiotemporal data require visual and analytical tools in order to improve the decision-making process. These tools should be intuitive so that little time is spent obtaining relevant conclusions from the analysis process, such as information on predictions, recurrence patterns, and clustering.

Visualization techniques are well known for improving the decision support process [1], once they take advantage on the human skills for quickly perceiving and interpreting visual patterns [2][3]. However, it has been argued that the visualization resources provided by most of the existing geographic-based applications are not enough for decision support systems when used solely [4].

Furthermore, spatiotemporal data impose serious challenges for analytics. First of all, due to the geographical space complexity, that requires human involvement and

his/hers sense of determination of space, place, and spatial relationships [5]. Secondly, due to the complexity of the temporal dimension. Time flows linearly, however, some events that occur over time may be periodically recurrent, with multiple cycles, forming hierarchical structures that overlap and interact with each other. Hence, temporal data analysis also requires human involvement [6].

It is necessary to perform analysis over data stored in heterogeneous and distributed data sources. In addition to the complex features of spatiotemporal data, the existence of heterogeneous sources results in an interoperability problem. Aiming at minimizing such problem, the Open Geospatial Consortium (OGC) [7] proposes standards for heterogeneous spatial databases connectivity services, such as the Web Map Service (WMS) and the Web Feature Service (WFS), broadly used by GIS (Geographic Information Systems) applications.

We propose in this paper a new approach, the GeoSTAT (Geographic Spatiotemporal Analysis Tool) system in order to address the problem of the lack of systems that may provide visualization and clustering techniques for large spatiotemporal datasets. GeoSTAT is a web-based environment that implements several spatiotemporal data visualization technique; interoperates with distributed data sources through OGC WMS and WFS services; and provides several data mining clustering algorithms proposed in the literature.

The main contribution of this work concerns the proposal of a visual approach for the analysis of spatiotemporal data that:

- Facilitates the exploration of the spatial and temporal dimensions;
- Provides interoperability and domain independence; and
- Integrates data mining algorithms into the visual analysis.

The remainder of the paper is organized as follows. Section 2 discusses related work. Section 3 focuses on the proposed environment. Section 4 addresses a case study to validate the proposed ideas. Finally, Section 5 concludes the paper and points out further work to be undertaken.

## II. RELATED WORK

There are many works based on spatial data visualization, but not many deal with the visualization of spatiotemporal data.

Reda et al. [8] present a tool that enables the visual exploration of the changes in dynamic social networks over time. It is an application focused on the domain of social networks visualization and it introduces a data structure based on a 3D cube for the spatiotemporal visualization.

Lu et al. [9] address a web-based system for the visualization of historical spatiotemporal data of the metropolitan area of Washington, D.C., in the United States. Apart from being an application based on a specific domain, their tool does not use georeferenced maps, dealing with the data visualization merely by several chart types.

He et al. [10] highlight another domain specific study. The authors developed a spatiotemporal data visualization system on the domain of oceans. It is a web-based system that uses 2D and 3D maps for the spatial visualization. The time visualization is static (has no animations), based on charts and triggered by user that chooses a region of interest in the map and a target time interval.

Chen et al. [11] implemented a tool that integrates several visualization techniques (GIS, self-organizing maps, hierarchical lists, periodical views, timeline views, etc.) for the criminal analysis domain. Although it is domain specific, this tool introduces interesting functionalities, like a time slider that allows time variations of the data over a georeferenced 2D map with basic interactivity tools such as pan and zoom. In spite of using several visualization techniques, the interface of the proposed application may get overloaded, making the user confused.

We have seen so far works that address spatiotemporal data visualization exploring several visualization techniques chosen according to specific domain. Next, we relate important works that propose several spatiotemporal visualization techniques.

Andrienko et al. [12] propose a framework based on the Self-Organizing Map (SOM) technique, combined with a number of interactive visualization techniques for the analysis of spatiotemporal data from two perspectives: spatial distributions that vary over time; and local time variation profiles distributed over time. This approach promises to be domain independent, gathering visualization techniques based on maps to enable the data analysis.

Compieta et al. [13] focus on issues related to the complexity of the manipulation, analysis and visualization of spatiotemporal data sets. The authors propose a spatiotemporal data mining system based on association rules and several techniques for the visualization and interpretation of georeferenced maps.

Andrienko et al. [14] argue that it is necessary to handle the time more effectively and list some characteristics that would be ideal for a good spatiotemporal data visual analysis system, such as treating and using both time and space characteristics, being visual, exploratory, scalable, collaborative, providing applicable methodologies for new

and big data sets, and providing mechanisms for gathering evidences.

Considering the related works that involve spatiotemporal data visualization, most of them address specific domain solutions and do not present flexibility on obtaining the data, forcing the user to use solely the data source provided by the application.

It is necessary to conceive a spatiotemporal data visualization and analytical tool that is, mainly: flexible, to enable the user to manipulate data obtained from information sources that and to execute spatial or temporal queries over these data, according to the analysis criteria; practical, in the sense of providing an intuitive interface, with resources that assist the visualization and analysis of both spatial (map resources) and temporal features (charts, map animations); and generic, by providing all those resources for users interested on any spatiotemporal analysis domain.

It is then necessary to use visual analytical tools for spatiotemporal data, together with data mining algorithms that will enable the discovery of implicit knowledge. This is the aim of the GeoSTAT system.

## III. THE GEOSTAT SYSTEM

This section introduces the GeoSTAT (Geographic Spatiotemporal Analysis Tool) system that implements the visualization and analysis of spatiotemporal data available on heterogeneous databases. We followed the guidance for good spatiotemporal visual analysis systems proposed by Andrienko et al. [14].

The GeoSTAT system is a web-based visualization tool, designed in three-tier architecture: Visualization, Control and Persistence. The first two tiers contain the core of our contributions. Figure 1 shows such architecture.

### A. The Visualization Tier

The visualization tier is responsible basically for the user interface. The GeoSTAT viewer was implemented based on the Google Maps API [15] and provides, in addition to the basic map interactivity functionalities (drag, pan, zoom, information and scale), options for alternating between base map types (map, satellite or terrain), and adding map layers. The viewer enables the visualization of several map layers, whether they are spatial or spatiotemporal, simultaneously, regardless of the data source. For each map layer, it is possible to apply an opacity level (transparency) to allow a better visualization of the spatial information. It is also possible to execute spatial and non-spatial queries over the visible map layers. Figure 2 shows the GeoSTAT visualization screenshot with one spatiotemporal layer added to the map.

Figure 2 shows the presence of temporal controllers, in addition to the spatial data manipulation tools already presented. For the spatiotemporal map layers, the viewer offers some components such as the Temporal Slider (see component 1 in Figure 2) that allows (see component 1 in Figure 2) that allows data visualization according to their timestamps and over several possibilities of distinct tempo-

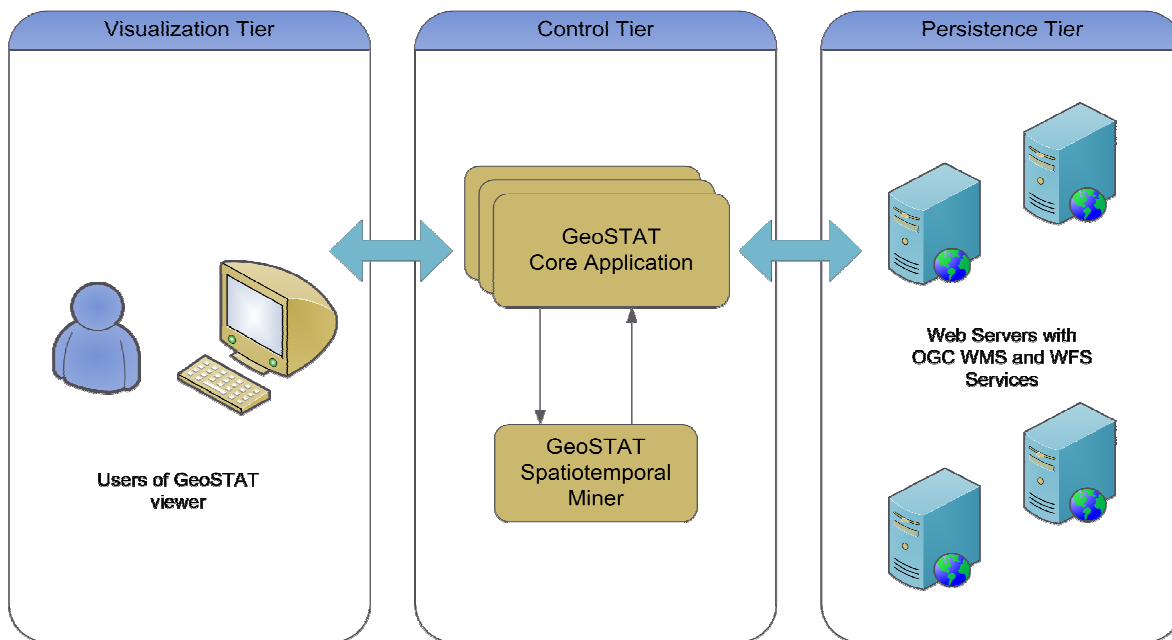


Figure 1. The GeoSTAT three-tier architecture.

-ral granularities, for example: day, month or year. It is also possible to apply temporal filters in order to reduce the number of data analyzed for relevant periods (moments or intervals), and to use interactive charts (see component 2 on Figure 2) to assist the analysis of the temporal distributions of the data.

A map layer is considered being spatiotemporal as from the moment of its inclusion in the application, the user indicates the temporal attribute of the layer.

Still at the visualization tier it is possible to execute and view the results of the clustering based spatiotemporal data mining on the spatiotemporal layers of the type POINT. Hence, the simultaneous visualization of a layer with real data and resulting data from the data mining processing may be achieved.

### B. The Control Tier

The control tier is where the user requests from the visualization layer are treated and executed. This tier implements the application logic and direct communication with the data services that offer the spatial or spatiotemporal layers. It is responsible, mainly, for the spatial and temporal queries, resulting from the use of the Temporal Slider, and for the execution of the spatiotemporal data mining.

Inside the control tier there is a module responsible for the spatiotemporal data mining. This module integrates the clustering algorithms found in the Weka data mining library [16] and is extensible for the addition of other algorithms of the same nature. To provide such extensibility, we developed a communication interface that can be easily implemented for new algorithms and we established a data input format based on spatiotemporal points; and a data output format for the spatiotemporal clusters. The input format chosen was the Comma-Separated Values (shortly

CSV), containing information on the spatiotemporal layer to be mined, such as the latitude, longitude and timestamp of the records. The output format chosen was XML (eXtensible Markup Language), containing relevant information over the generated clusters. An example of a generated cluster on the output format of the data mining module is presented in Code 1.

According to Code 1, each element of the type "cluster" is basically composed of an identifier, the number of instances (records) grouped (representing its density) and the cluster's spatial and temporal elements. The spatial elements enable the visualization of the cluster on georeferenced circle format, with radius and center point clearly defined. The temporal elements specify the temporal granularity and the value.

Code 1. Snippet of the XML file that represents an example of a cluster generated by the data mining module.

```

<cluster>
  <id>1</id>
  <instances>144</instances>
  <spatial>
    <latCentroid>-8.253</latCentroid>
    <lonCentroid>-36.964</lonCentroid>
    <radius>0.14567</radius>
  </spatial>
  <temporal>
    <granularity>YEAR</granularity>
    <value>2010</value>
  </temporal>
</cluster>
    
```

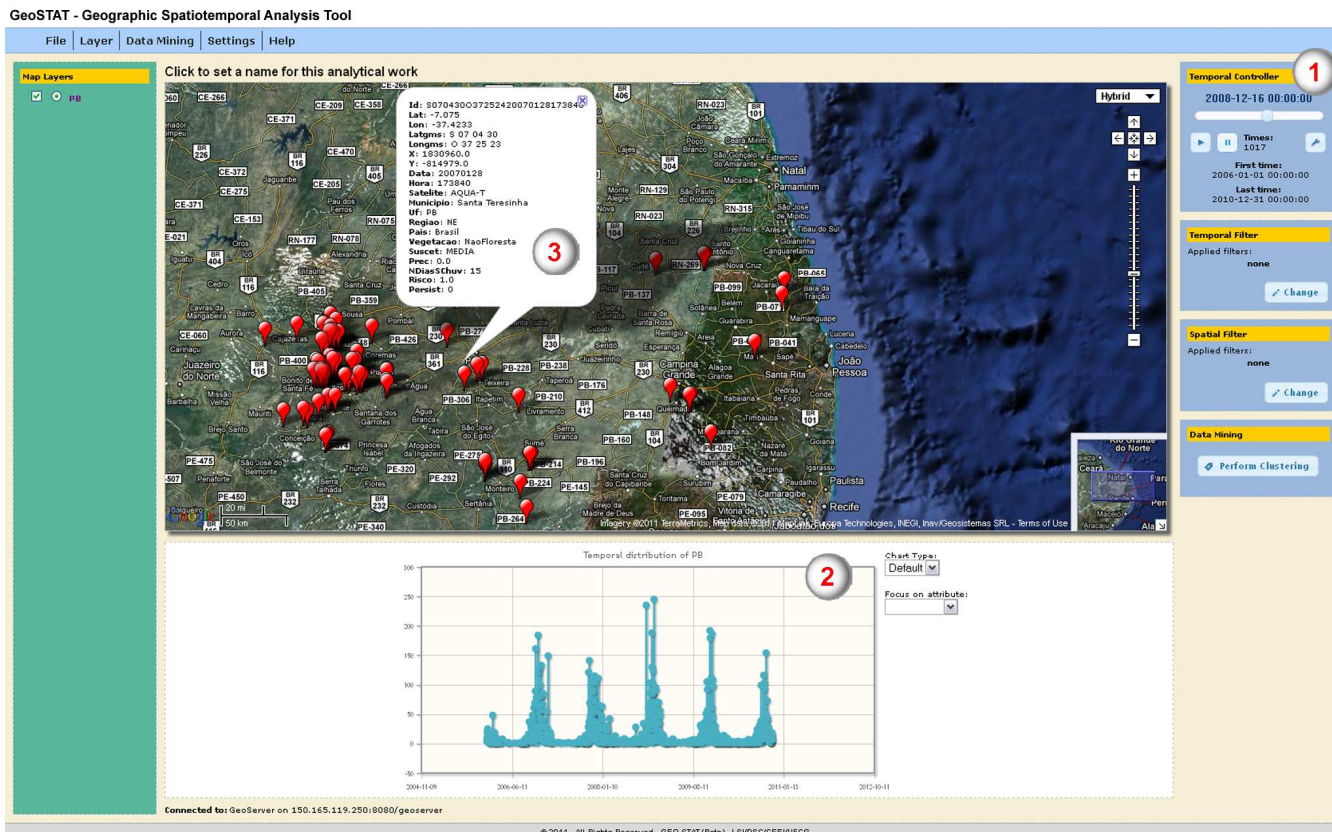


Figure 2. The GeoSTAT environment visualization screenshot with a spatiotemporal map layer.

### C. The Persistence Tier

The persistence tier contains GeoServer map server that implements the OGC WMS and WFS services [17], ensuring the spatial DBMS (Database Management System) interoperability. The servers used by the users through GeoSTAT may be easily connected, through the information of the service URL and setting an alias to better identify it. With an established connection, the user has the option to perform map overlay of all the layers available from the OGC connected services.

## IV. CASE STUDY

This section presents a case study in order to validate the solution proposed in this paper.

To explore the functionalities offered by GeoSTAT, we set up a web server with GeoServer version 2.0.3, that implements the WMS and WFS services. We also used the PostgreSQL DBMS version 9.0.4, with the PostGIS spatial extension version 1.5.

The spatiotemporal data set was obtained from the Brazilian National Institute for Space Research (INPE) [18], resulting in 17,418 records.

The data set contains records of fire events detected by satellites in the state of Paraiba, located in the Northeast region of Brazil, during a period of five years (2006-2010), resulting in 17,418 records. We used spatial data obtained from the Brazilian Institute of Geography and Statistics

(IBGE) to visualize the vector layers of the states of the Northeast of Brazil (9 polygons), and those from the cities of the Paraiba state (223 polygons),

Through GeoSTAT we may set up a data connection to the web server and obtain a list of available layers for inclusion, visualization and analysis. Figure 3 shows a screenshot for adding a new data connection. GeoSTAT may connect to any map servers that implement OGC WMS and WFS services. In the case study we used the GeoServer map server.



Figure 3. Adding a spatial data connection.

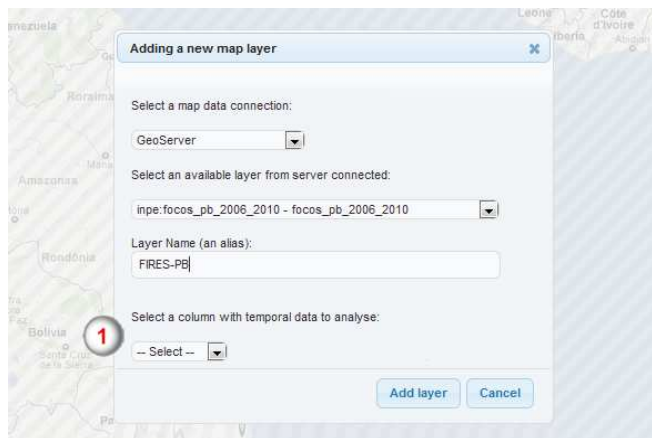


Figure 4. Adding a spatiotemporal map layer.

Figure 4 shows the screenshot of adding map layers. The layers available from the server are obtained by the *GetCapabilities* request, specified in the OGC WMS service. If a layer contains spatiotemporal data, the user has the option to select the attribute that contains the temporal dimension (see item 1 in Figure 4). The layer attributes are acquired using the *DescribeFeatureType* request, specified by the OGC WFS service.

The selection of the temporal attribute is an essential step to perform spatiotemporal analysis. By specifying this attribute, GeoSTAT provide temporal components to the user. Otherwise the system will treat the attribute as spatial and will make available only the spatial API.

In our experiment, we used the following layers: States (spatial), Cities (spatial) and Fires (spatiotemporal). Figure 5 shows the GeoSTAT viewer with these layers visible in the map. Map layers are obtained from the servers through the *GetMap* request, specified by the WMS service. We may control the visibility of layers (see item 1 in Figure 5). The "checkbox" corresponds to the control of the visibility of the layer, whereas the "radio button" to the active layer control (see item 3 in Figure 2).

Figure 6 shows the information component, which is enabled by clicking on the visible layer, e.g., the FIRES layer. This is done through the WMS *GetFeatureInfo* requisition. The layer opacity may be changed by clicking on the layer name.

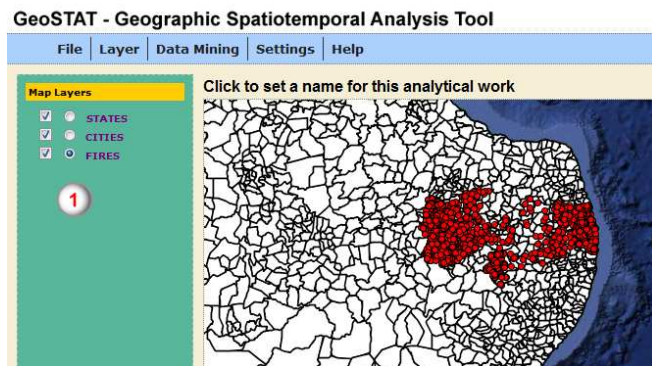


Figure 5. GeoSTAT viewer with three layers visible in the map.

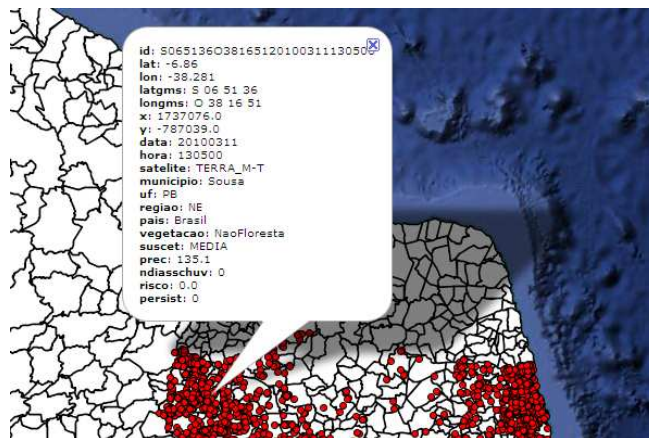


Figure 6. An example of the information component.

After presenting the "Fires" spatiotemporal layer in the map, GeoSTAT displays a chart area with the distribution of such events over time (see item 2 in Figure 2). In the chart area, it is possible, for example, to classify the distribution chart using nonspatial and nontemporal attributes. In our experiment, we used the vegetation attribute to generate a new distribution chart showing types of vegetation over time (see Figure 7). Then the temporal slider is enabled so that the user may press the "play button" to start a map animation over time (see item 1 in Figure 2).

Figure 8 presents an example of the temporal animation component. There are interface components such as play and pause; and a temporal slider. The objects shown in the map are exhibited according to the specific time being played.

As the "States" and "Cities" layers include all of the Brazilian territory and we would like to focus on the region of fires analysis (Northeast region), we used the spatial filter to show on the map only the state of Paraíba and its 223 cities. The use of spatial filters is only possible due to the *GetFeature* request specified by the OGC WFS service, which enables to execute queries on the spatial layers. Then, the user may be interested in analyzing the detected fire events where the type of vegetation was "NoForest" and during the year of 2008. By analyzing the dynamic chart, we can see which type of vegetation registered the highest concentration of such events in that time period. To perform this operation, we first apply a spatial filter on the layer, in order to display in the map only the records associated with the vegetation type "NoForest", and then we

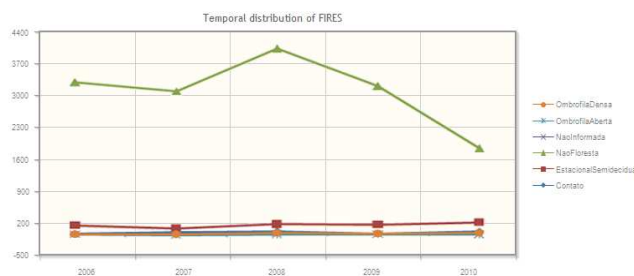


Figure 7. Fire distribution chart over time, classified by type of vegetation.

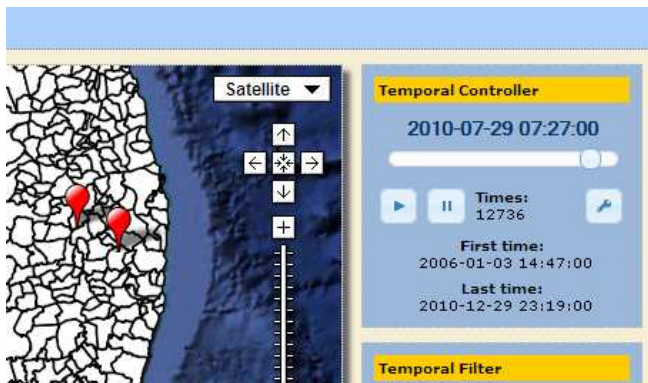


Figure 8. Using the temporal controller

apply a temporal filter for the time period between "2008-01-01" and "2008-12-31". That way, we reduced the result set to 4,010 records of fires to be analyzed. The results may be seen in the map shown in Figure 9a.

Finally, we execute the spatiotemporal data mining algorithms to the fire dataset aiming to find out relevant implicit patterns. For instance, we may detect that during a certain period of the year, a given type of vegetation is severally affected by fires in a specific geographic region (cluster). This knowledge may help decision makers to study fire defense and emergency services planning for future occurrences. We use the month granularity, through the density-based DBScan clustering algorithm [19], provided by GeoSTAT, with parameters  $\epsilon = 0.1$ ,  $\text{minPoints} = 2$  and  $\text{distance-type} = \text{'Euclidian distance'}$ . The input values for the DBScan algorithm were established empirically. The resulting clusters can be seen in Figure 9b.

DBScan [19] is a density based cluster algorithm that groups fire spots using spatiotemporal neighborhood. Hence, the fires occurred in a given time interval and within a distance specified in the epsilon parameter will constitute a cluster. The "minPoint" parameter specifies how many neighbor points are necessary to obtain a cluster. Finally, the distance type parameter specifies the distance metric to be used; in our case we are using the Euclidian distance.

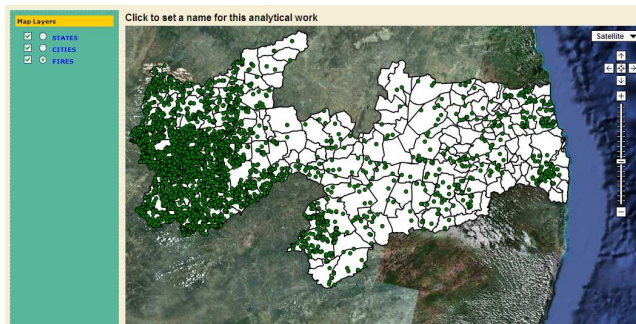
We validate our approach in a real scenario, and we could the usability and effectiveness of the proposed system.

## V. CONCLUSION AND FUTURE WORK

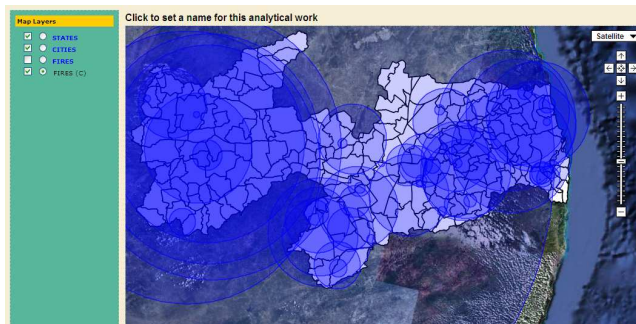
We addressed the question of spatiotemporal visualization and analysis, highlighting its utility in the presence of large data sets. We also pointed out the main limitations found in the related work.

We proposed a spatiotemporal visualization and analysis environment aiming to minimize the main limitations found, regarding the flexibility of the data sources, practicality of the analysis process and generalization of the domain. Our case study has shown that the proposed solution fulfills the requirements established, being valid not only in the domain adopted for the case study, but also on any other domains.

We built a DBMS independent environment, following the OGC guidelines specifying the WMS and WFS services that allow the interoperability between several data sources



(a)



(b)

Figure 9. Case Study results: (a) 4,010 fires where type of vegetation is equals to "NoForest" (transactional data) during the year of 2008. (b) density spatiotemporal clusters generated of these 4,010 fires (mined data).

with the use of simultaneous map layers regardless of their origins, in a transparent way for the user.

As further work, we plan to improve GeoSTAT by incorporating a 3D viewer for spatiotemporal data. Also, the addition of the collaborative analysis concept, allowing several analysts to work in a shared and evolutionary manner is another interesting future work.

## REFERENCES

- [1] W. Johnston, "Model visualization," in *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002, pp. 223–227.
- [2] I. Kopanakis and B. Theodoulidis, "Visual data mining modeling techniques for the visualization of mining outcomes," *Journal of Visual Languages and Computing*, vol. 14, no. 6, 2003, pp. 543–589.
- [3] N. Andrienko, G. Andrienko, and P. Gatalsky, "Exploratory spatiotemporal visualization: an analytical review," *Journal of Visual Languages and Computing*, special issue on Visual Data Mining, vol. 14, no. 6, 2003, pp. 503–541.
- [4] Y. Bédard, T. Merrett, and J. Han, "Fundamentals of spatial data warehousing for geographic knowledge discovery," in H. J. Miller and J. Han (eds.) *Geographic Data Mining and Knowledge Discovery*, London: Taylor and Francis, 2001, pp. 53-73.
- [5] G. Andrienko, N. Andrienko, J. Dykes, S. I. Fabrikant, and M. Wachowicz, "Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research," in *Information Visualization*, vol. 7, no. 3, 2008, pp. 173–180, doi:10.1057/ivs.2008.23.
- [6] D. J. Pequet, "Representations of space and time", The Guilford Press, 2002.
- [7] A. T. Kralidis, "Geospatial Open Source and Open Standards Convergences," in G. B. Hall and M. G. Leahy (Eds.) *Open Source Approaches in Spatial Data Handling*, Berlin: Springer, 2008, pp. 1–20.

- [8] K. Reda, C. Tantipathananandh, T. Berger-Wolf, J. Leigh, and A. Johnson, "SocioScape - a Tool for Interactive Exploration of Spatiotemporal Group Dynamics in Social Networks," in Proceedings of the IEEE Information Visualization Conference (INFOVIS '09), Atlantic City, New Jersey, 2009.
- [9] C.-T. Lu, A. P. Boedihardjo, and J. Zheng, "Towards an Advanced Spatiotemporal Visualization System for the Metropolitan Washington D.C." in 5th International Visualization in Transportation Symposium and Workshop, 2006, pages: 6.
- [10] H. Yawen, S. Fenzhen, D. Yunyan and X. Rulin, "Web-based visualization of marine environment data," in Chinese Journal of Oceanology and Limnology, vol. 28, no. 5, Science Press, co-published with Springer-Verlag GmbH, 2010, pp. 1086–1094.
- [11] H. Chen, H. Atabakhsh, T. Petersen, J. Schroeder, T. Buetow, L. Chaboya, C. O'Toole, M. Chau, T. Cushna, D. Casey, and Z. Huang, "COPLINK: Visualization for Crime Analysis," in Proceedings of The National Conference on Digital Government Research, 2003, pp. 1–6.
- [12] G. Andrienko, N. Andrienko, S. Bremm, T. Schreck, T. V. Landesberger, P. Bak, and D. Keim, "Space-in-Time and Time-in-Space Self-Organizing Maps for Exploring Spatiotemporal Patterns," in Computer Graphics Forum, vol. 29, no. 3, 2010, pp. 913–922.
- [13] P. Compieta, S. D. Martino, M. Bertolotto, F. Ferrucci, and T. Kechadi, "Exploratory spatiotemporal data mining and visualization," in Journal of Visual Languages & Computing, vol. 18, no. 3, 2007, pp. 255–279.
- [14] G. Andrienko, N. Andrienko, U. Demsarb, D. Dransch, J. Dykes, S. I. Fabrikant, M. Jernf, M.-H. Kraak, H. Schumann, and C. Tominskih, "Space, time and visual analytics," in International Journal of Geographical Information Science, vol. 24, no. 10, 2010, pp. 1577–1600.
- [15] Google Inc, "Google Maps Javascript API V2 implementation reference documentation," 2010, available from: <http://code.google.com/apis/maps/documentation/javascript/v2/reference.html> 25.11.2011.
- [16] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," in SIGKDD Explorations, vol. 11, no. 1, 2009, pp. 10–18.
- [17] Open Geospatial Consortium, "GeoServer - a Java-based software server to view and edit geospatial data," 2008, available from: <http://geoserver.org/display/GEOS/What+is+Geoserver> 25.11.2011.
- [18] INPE, "Vegetation Fires - Fire Monitoring," in Brazilian National Institute for Space Research, 2011, available from: [http://sigma.cptec.inpe.br/queimadas/index\\_in.php](http://sigma.cptec.inpe.br/queimadas/index_in.php) 25.11.2011.
- [19] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), 1996, pp. 226-231.