Finding Spatio-temporal Patterns in Multidimensional Data Streams

Santiago A. Nunes[1], Luciana A. S. Romani[2], Ana M. H. Avila[3], Priscila P. Coltri[3], Agma J. M. Traina[1], Elaine P. M. Sousa[1]

[1] University of São Paulo at São Carlos, Brazil
{santiago, agma, parros}@icmc.usp.br
[2] Embrapa Agricultural Informatics - Campinas - Brazil
luciana.romani@embrapa.br
[3] Cepagri - University of Campinas - Brazil
{avila, pcoltri}@cpa.unicamp.br

Abstract. In the last few decades, advances in data acquisition technology have contributed to generation of huge volumes of data in diverse application areas, creating new research challenges in knowledge discovery. The analysis of these data has become an important task in several domains such as sensor networks, web-logs, financial transactions and climate change monitoring. In this article, we propose the Spatio-Temporal Behavior Meter (STB-meter) method to identify spatio-temporal patterns in multidimensional evolving data streams. Our approach combines a multi-resolution hierarchical structure to deal with spatial information with fractal-based analysis to monitor non spatial information of the multidimensional data stream. Experimental evaluation on real climate data shows that our method allows finding relevant spatio-temporal patterns in evolving data at different spatial and temporal resolutions and therefore it can be a useful tool to assist domain specialists in climate change researches.

Categories and Subject Descriptors: H.2 [**Database Management**]: Miscellaneous

Keywords: fractals, multi-resolution spatial structure, spatio-temporal data

1. INTRODUCTION

In this century scientists from different fields of knowledge have the challenge of dealing with big data. One reason is the increasing number of data generation environments spatially distributed and interrelated. Examples of these environments include different kinds of sensor networks connecting remote sensors, mobiles, RFID, GPS and so on.

Another motivating factor is the improving quality of technological equipments that allow more and more data to be generated or captured in real time. This continuous flow of a very large data volume can quickly deplete storage capacity or overload systems. In general, much of these data might be analyzed / monitored continuously without storage. In this scenario, computational techniques that have not been designed to deal with continuous data efficiently may quickly become obsolete.

Therefore, the new generation of algorithms for knowledge discovery should consider new issues, such as space, time and data structure to deal with huge volumes of data. Additionally, the formulation and submission of responses must be faster, as well. The number of mobile devices grows very quickly and information is accessed anytime and anywhere. Also, sensor networks have been deployed to assist

The authors thank Embrapa, FAPESP, Microsoft Research, CNPq and Capes for the financial support, and Agritempo for the climate data used in this work.

in monitoring different aspects of modern life, such as climate, vehicles traffic, and so on.

A sensor network can involve a large number of different devices, depending on the complexity and extent of territory that it is intended to monitor. Therefore, the computational challenge becomes even greater since it becomes necessary to combine the spatial information to the increasing flow of data.

The spatio-temporal analysis of data streams can handle a continuous data stream, considering the data reading, timestamp and also the geographic position where the sensor (data source) is located. The algorithms for mining spatio-temporal data streams must run efficiently in order to properly detect recent patterns considering the georeferenced data, representing a new challenge in the knowledge discovery area [Gama 2010]. As an example, studies on climate change have imposed several challenges to scientific and governmental communities. Food security and sustainable development are certainly the most important. Additionally, terabytes of data are generated during each execution of climate models, depending on the number of parameters employed, which commonly goes up to a few tens of parameters such as temperature, humidity, wind direction and intensity, among others. According to the last IPCC (Intergovernmental Panel on Climate Change) report, climate change could cause a migration of crops adapted to a specific area to other regions to compensate for alterations on climate conditions. As climate scenarios show different changes in the future, consequences for agricultural production may have greater or smaller impact in different regions.

In this context, we propose the *Spatio-Temporal Behavior meter* (STB-meter), a method based on a hierarchical data structure to represent multidimensional data streams in order to allow a spatio-temporal assessment. The STB-meter's approach integrates representation of spatial information in a hierarchical data structure and the measure of the correlation fractal dimension D_2 to detect spatio-temporal patterns and changes in the stream's behavior.

We have conducted experiments on real climate data, such that multiple georeferenced climate time series are represented as a multidimensional data stream - each time series defines an attribute of the stream. Therefore, it is possible to integrate multiple climate variables in a unified analysis process, aiming to detect spatio-temporal patterns, behavior changes and extremes. Results have shown that the proposed method is effective to detect spatio-temporal variations in data behavior. In particular, experiments performed on data from meteorological sensors located in the state of São Paulo show that STB-meter allows a spatio-temporal analysis of real data and therefore it is also useful in applied sciences.

This article is organized as follows: Section 2 presents background concepts and related work. Section 3 describes the *Spatio-Temporal Behavior meter* method, the approach we propose to conduct a spatio-temporal analysis of multidimensional data streams. Section 4 details experimental results on real climate data and Section 5 presents final remarks.

2. BACKGROUND AND RELATED WORK

The method we propose in this work integrates fractal-based analysis with a multi-resolution, hierarchical data structure in order to support spatio-temporal analysis of evolving data streams. Thus, this section provides an overview of concepts from the Fractal Theory applied to data analysis, spatio-temporal issues and related work.

A fractal is a self-similar object, i.e., it presents roughly the same characteristics over a large range of scales [Schroeder 1991]. A well-known example of fractal is the Sierpinski Triangle, illustrated in Figure 1a. Fractal behavior is also observed in nature and real datasets, as illustrated in Figures 1b and c. A real dataset exhibiting fractal behavior is exactly or statistically self-similar, such that parts of any size of the data present the same general characteristics of the whole dataset. In fact, Faloutsos and Kamel [1994] show that most of the real datasets exhibit fractal behavior.

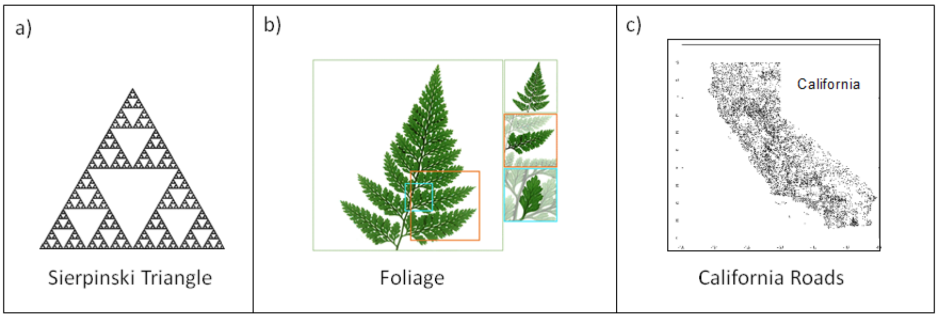

Fig. 1. Examples of Fractals: a) geometric fractal; b) fractal in nature; c) real fractal dataset.

The fractal dimension, in particular the Correlation Fractal Dimension D_2, is a useful tool for data analysis, as it provides an estimate of the intrinsic dimension D of real datasets. The intrinsic dimension gives the dimensionality of the object represented by the data regardless of the dimension E of the space in which it is embedded [Faloutsos and Kamel 1994; Traina Jr. et al. 2005]. For instance, a set of points defining a line embedded in a three-dimensional space ($E = 3$) has all its attributes correlated, resulting in $D = 1$.

A well-known approach to measure the fractal dimension of datasets embedded in E-dimensional spaces is the *Box-Counting* method [Schroeder 1991], which defines D_2 as presented in Equation 1:

$$D_2 \equiv \frac{\partial log(\sum_i C_{r,i}^2)}{\partial log(r)} \qquad r \in [r_1, r_2] \tag{1}$$

where r is the side of the cells in a (hyper) cubic grid that divides the address space of the dataset and $C_{r,i}$ is the count of points in the ith cell.

Based on the Box-Counting method, Traina et al. [2000] proposed a $O(N)$ algorithm (N is the number of elements in the dataset) to compute D_2. The main strategy of the algorithm is the construction of a hierarchical data structure (*counting tree*) to map a multi-resolution hyper-grid dividing the address space of the dataset. Each level of this hyper-grid has a radius r which is a fraction of the previous level, i.e., in the first level the grid has a radius r, in the second level the grid has radius of $r/2$ and so on. The counting of incident points is accomplished in each cell of the hyper-grid as we can see in Figure 2. The *counting tree* supports fast counting of points for different values of r (see the work of Traina et al. [2000] for details). Thus, D_2 can be a useful tool to estimate the intrinsic dimension D of real datasets with feasible computational cost.

Concepts from the Fractal Theory have been applied to several tasks in data mining and data analysis, such as selectivity estimation [Böhm 2000; Faloutsos et al. 2000; Baioco et al. 2007], clustering [Barbará and Chen 2010; Cordeiro et al. 2013], time series forecasting [Chakrabarti and Faloutsos 2002], correlation detection [de Sousa et al. 2007], data distribution analysis [Traina Jr. et al. 2005], among others.

The information of intrinsic behavior provided by the fractal dimension can also be applied to detect temporal patterns and changes in evolving data streams. Essentially, the idea is to continuously measure the fractal dimension of the data stream over time in order to monitor its evolving behavior. Thus, significant variations in successive measures of the fractal dimension can indicate changes in the intrinsic characteristics of the data and temporal patterns.

Sousa et al. [2007] propose a technique and the algorithm *SID-meter* to continuously measure the fractal dimension D_2 over time and track behavior changes of evolving data streams. The *SID-meter*'s

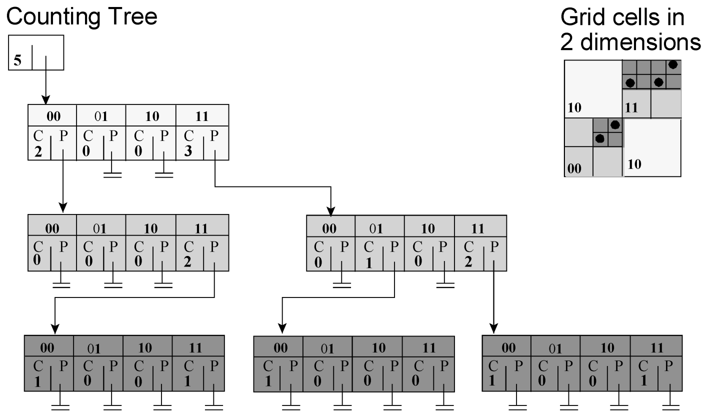

Fig. 2. Five points of a bi-dimensional dataset represented in a bi-dimensional grid and the corresponding *counting tree*.

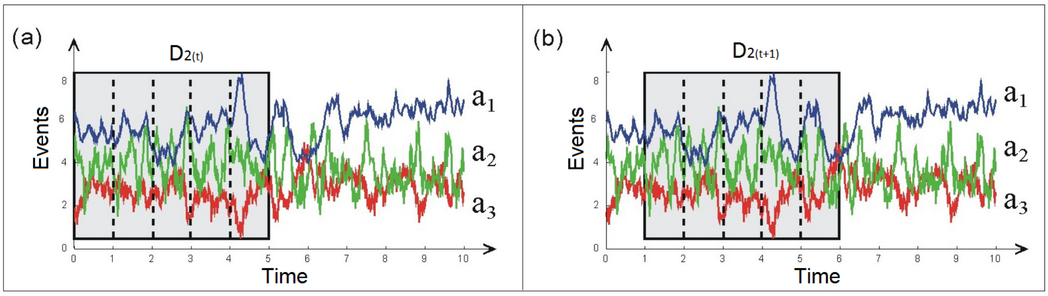

Fig. 3. A sliding window over a three-dimensional data stream used to measure fractal dimension at time t and time $t+1$, respectively: a) $D_{2_{(t)}}$; and b) $D_{2_{(t+1)}}$

approach deals with a data stream as a potentially unbounded, implicitly ordered sequence of events $< e_1, e_2, ... >$, such that each event is represented by an array of E measures (attributes). A sliding window bounds recent, successive events to be considered into the calculation of D_2. As new events come, oldest events are forgotten. Therefore, the value of D_2 is continuously computed for the events inside the window and updated whenever new events are available. The window is divided into n_c periods where each period is defined by a predetermined number of events or units of time (n_i). Hence, $n_c \times n_i$ gives the size of the window and n_i determines its update interval. Both n_c and n_i are user-defined parameters. Figure 3 illustrates a sliding window over a three-dimensional data stream - attributes a_1,a_2 and a_3 - and the fractal dimension D_2 measured at times t and $t+1$. Notice that the window has five periods ($n_c = 5$), each one corresponding to a unit of time $n_i = 1$.

SID-meter is also based on a *counting tree* tuned to support important requirements of a data stream environment, mainly: single pass on data, events processed (counted) only once and updated responses.

The fractal-based approach, as discussed herein, can be applied to temporal analysis of data streams. However, in several data stream environments, data includes spatial information related to its generators. The spatial characteristics and relationships may be meaningful for data analysis and mining tasks. Moreover, by dealing with spatial properties in addition to temporal information it is possible to conduct data stream analysis from the spatio-temporal perspective. In this context, some work has been done, for instance, on stream modeling and trend mining [Meng and Dunham 2006], clustering on sensor networks [Yoon and Shahabi 2007] and spatio-temporal continuous query processing [Mokbel and Aref 2008].

Table I. Example of a sequence of events that compose a data stream.

Events	Attributes
e_1	{1,1,4,8}
e_2	{4,4,5,8}
e_3	{7,7,8,2}
e_4	{8,6,6,3}
e_5	{7,5,4,5}

In this article, we propose a technique for spatio-temporal analysis of data streams which combines the temporal information provided by the fractal-based approach with spatial indexing provided by a hierarchical data structure [Samet 1984].

Hierarchical data structures are of major importance as a tool in several computational techniques, from image processing to geographic information systems. These structures are based on recursive decomposition of the data space and therefore allow focus on subsets of the original dataset. Several research papers have proposed indexing structures to represent points in space, regions, curves, surfaces and volumes (see the work of Gaede and Günther [1998] and Böhm et al. [2001] for surveys). In this work we have used a quadtree like structure as our hierarchical data structure, but it is worthy to highlight that the proposed method can be implemented with any hierarchical data structure to represent spatial attributes.

3. THE SPATIO-TEMPORAL BEHAVIOR METER

The main idea of the STB-meter is to continuously measure the fractal dimension D_2 of multidimensional data streams considering the spatial aspects of the data. For this purpose, we developed a data structure called STB-tree that allows the calculation of fractal dimension for different subsets of data based on their spatial information.

The STB-tree is built by associating a counting tree to a hierarchical data structure. In this work, we have used a quadtree like structure extended to any dimensionality.

The building process of the STB-tree requires prior knowledge about which attributes are the spatial attributes, as they are used to construct the hierarchical data structure. The non-spatial attributes are used to construct a counting tree, following an approach similar to the method SID-meter aforementioned.

The STB-tree integrates the hierarchical data structure with a counting tree modified to attend an indexing that depends on the subregion of the hierarchical data structure to which it is associated as well as the addressing defined by the non-spatial attributes. In other words, each level contains a set of subregions of the hierarchical data structure and each of these subregions contains a set of cells of the counting tree.

The cell indexing of the counting tree depends on the subregion of the hierarchical data structure to which they are related as well as their non-spatial attributes. In the same manner, indexing each subregion of the hierarchical data structure depends on the counting tree cell to which it is associated as well as its spatial attributes.

For illustration purposes, Table I presents a dataset containing 5 events where the first two attributes represent spatial information while the other two contain non-spatial information.

The STB-Tree is built according to the following steps:

(1) First define which attributes are spatial and non-spatial
(2) To each new event of data stream, split spatial attributes from non-spatial attributes

(3) Use spatial attributes to identify the subregion of the space represented by hierarchical structure
(4) In this subregion of the hierarchical structure use the non-spatial information to identify the counting tree cell corresponding to the subregion of the hierarchical structure
(5) Increment the cell's counter
(6) Use the spatial information to identify the hierarchical data structure subregion corresponding to the counting tree cell
(7) Repeat this process from step 4 recursively until the desired depth is reached in the STB-Tree

Figure 4 shows a STB-tree with depth 3 for the events shown in Table I. We can see the representation of spatial attributes on a quadtree as well as the representation of non-spatial attributes. For each level in the quadtree, one cell list is created in order to identify the position of the event in the counting tree of non-spatial attributes.

Thus, as it can be seen in Figure 4, the event e_1 showed in Table I has its spatial information defined in the values {1,1} and its non-spatial information defined in the values {4,8}. With the spatial information, initially the subregion 01 of the hierarchical structure is found. The algorithm then searches for the corresponding cell in the counting tree associated to that subregion. In Figure 4 we can verify that the counting tree cell is addressed in the index 00.

Analogously, the spatial information of event e_5 ({7,5}) is addressed in the subregion 10 of the hierarchical structure, and its non-spatial information ({4,5}) is also addressed in a cell 00 of the counting tree. However, as the spatial location of event e_5 differs from the location of event e_1, its contribution in the counting tree occurs in the cell 00 that is related to subregion 10 of the hierarchical structure.

This process is performed for all events in the dataset to be analyzed. Whenever a new event is inserted into the STB-Tree, the appropriate counter C is incremented, in order to store a count of events considering the spatial and non-spatial aspects of the dataset. To perform the calculation of fractal dimension of a particular region of the hierarchical structure, Equation 1 is used with the counting of the nodes of the STB-tree. Through STB-Tree it is possible to perform the counting of hierarchical subregions by accomplishing only one reading of each event.

Once data streams are potentially infinite and due to their evolving nature, we have used an approach based on sliding windows. The STB-tree has thus been extended to support a continuous analysis of multidimensional data streams, following the SID-meter strategy of setting parameters n_c and n_i to respectively define the number of counting periods and the number of events (or units of time) per each counting period.

The parameters n_c and n_i determine the temporal granularity used in the analysis, and therefore are strongly dependent on the type of analysis to be performed. For instance, if the analysis focuses on the events that occur monthly, an oversized window can smooth changes in the monthly behavior. On the other hand, if the size of the window is underestimated then behavior changes may not be detected. Previous work [Nunes et al. 2011] shows that windows of different sizes detect different events, such as El Niño and La Niña.

In order to enable a more complex analysis of the data, the sliding window based approach was extended such that parameters n_c and n_i receive a range of values. Based on these ranges, the minimum and maximum number of counting periods are computed so it is possible calculate the fractal dimension for any combination of n_c and n_i in the ranges defined by the user.

For this purpose, the STB-tree has also been extended so that each counter C uses an array to store the counts for each counting period. When the number n_i of events arrive, the counters corresponding to n_i oldest events are discarded and start to consider only the n_i new events.

Figure 5 illustrates the extended node of the STB-tree, with n_c ranging from 1 to 3 and n_i ranging

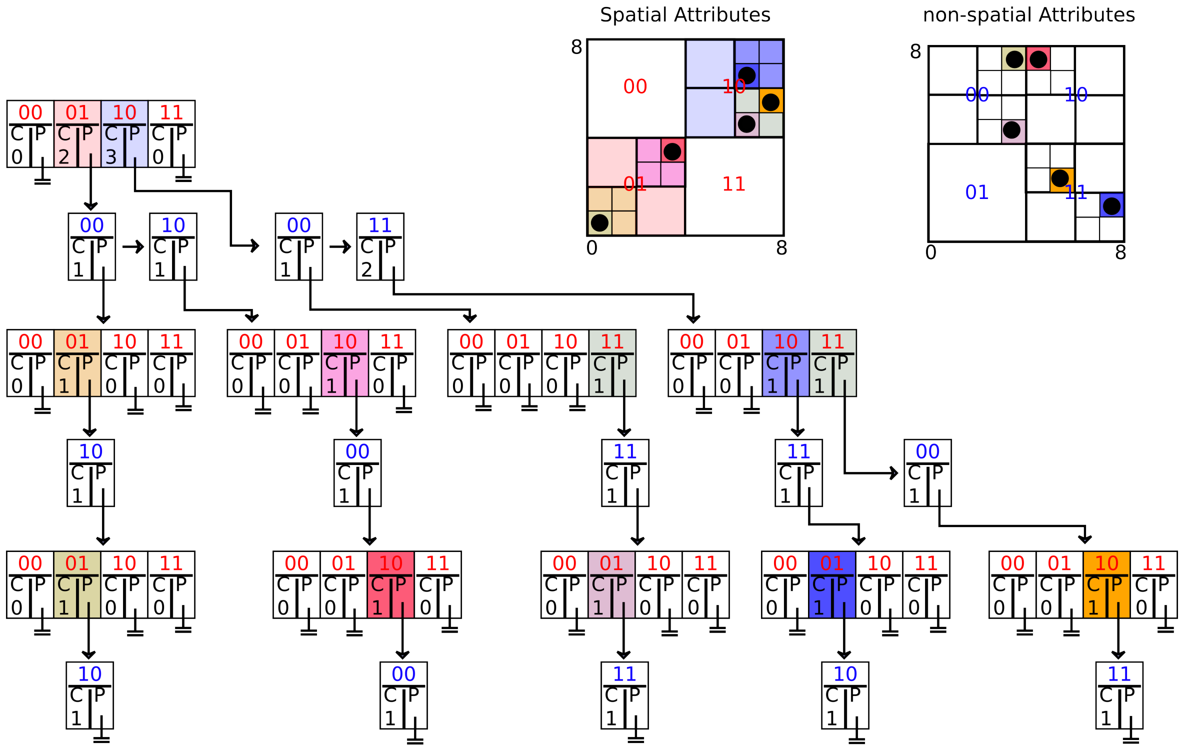

Fig. 4. STB-tree corresponding to the events shown in Table I.

from 30 to 120. In this example, it is possible to compute the fractal dimension for any combination of $n_c \subset \{1, 2, 3\}$ and $n_i \subset \{30, 60, 90, 120\}$. For instance, given $n_c = 2$, $n_i = 60$ and the current n_c counter positioned at 7, the count on the node will consider $C[4], C[5], C[6]$ and $C[7]$ to calculate the fractal dimension.

By using the STB-tree structure, the STB-meter method was developed to allow the continuous calculation of fractal dimension on different subsets of data. This method allows an analysis in multiple spatial resolutions as well as different temporal granularities, performing only one reading of each event in the data stream.

The output of the method are values of the fractal dimension D_2 calculated over time for each subregion represented in the hierarchical data structure, from the smallest subregions until it encompasses all the points represented. Analysis of these results allows to identify behavior changes in each subregion of the space and to find patterns, differences and similarities among the subregions over time.

The combined analysis of multidimensional data streams considering spatial and temporal aspects of the data set can be useful in detecting patterns and extreme events occurring in different regions

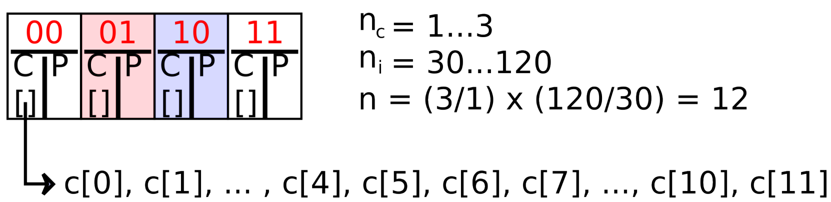

Fig. 5. STB-Tree Node extended to support sliding windows.

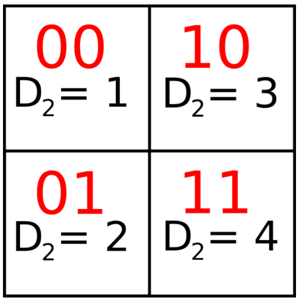

Fig. 6. Correlation of synthetic data regions.

and in different situations.

4. EXPERIMENTAL RESULTS

To evaluate the STB-meter method, experiments were performed using synthetic data streams constructed so that each region of the data set presents a specific known behavior. This set of synthetic data was projected to assess whether the STB-Meter can calculate correctly the fractal dimension of different subregions in space. Thus, each subregion of space in this dataset contains non-spatial attributes correlated in order to detect different values of correlation in each region.

Another goal of this synthetic set is performed tests in multiple levels in order to verify it is possible find behavior different from general behavior of dataset. For instance, a region presents a specific fractal dimension, and subregions which compose it presents a fractal dimension that differs it from that.

Figure 6 illustrates the arrangement of built synthetic dataset containing two spatial attributes and four non-spatial attributes.

This dataset contains 36000 events such that in region **00** all non-spatial attributes are correlated, then the fractal dimension expected to this region is $D_2 = 1$.

In the region **01** spatial attributes are divided into two subregions as illustrated in Figure 7, where subregion 00 contains pairs of correlated attributes, generating a fractal dimension $D_2 = 2$ and subregion **11** contains attributes without any correlation, with a fractal dimension $D_2 = 4$.

In region **10** only two non-spatial attributes are correlated, therefore, fractal dimension of this region is $D_2 = 3$ and in region **11** no non-spatial attribute is correlated, being $D_2 = 4$.

Over this dataset the STB-meter method was parameterized using windows divided into 10 counting periods where each period contains 2,400 events, i.e. $n_c = 10$ and $n_i = 2,400$, and uses a tree with deep equal to 21, i.e., $R = 21$. Results are illustrated in Figure 8 where it is possible to identify the different behavior of each region and how each region differently adds information regarding the entire dataset. The fractal dimension correlation calculated by the STB-meter method corresponds to the expected correlation for each region of the synthetic data set as presented in table II.

Table II. Arrangement of synthetic dataset in each region.

Spatial Index	D_2
00	1
01	2
10	3
11	4

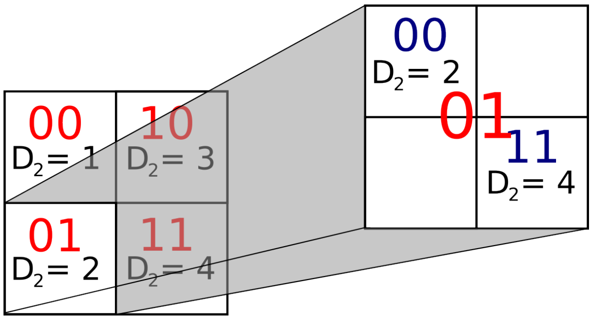

Fig. 7. Correlation of subregions of the region **01**.

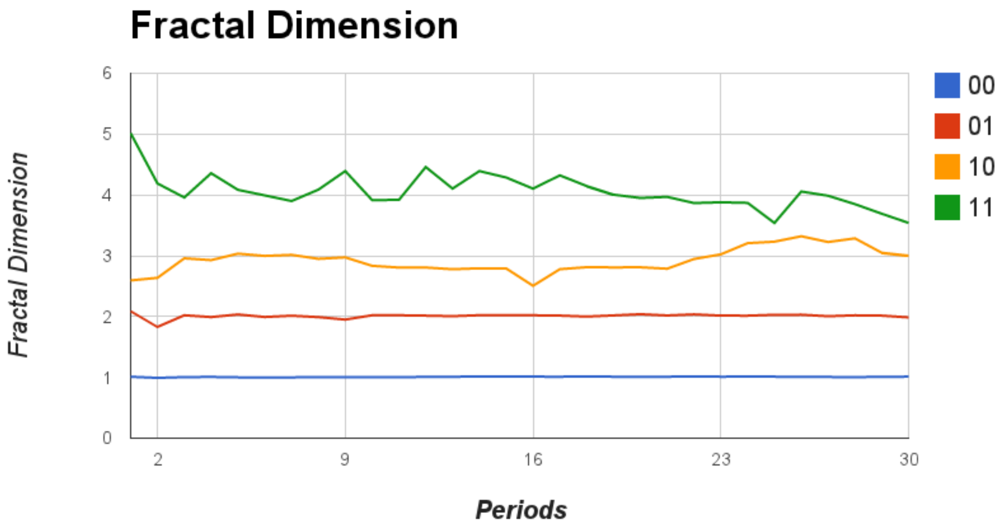

Fig. 8. Fractal dimension of synthetic data regions.

Similarly, region **01** can be divided into four subregions, as shown in Figure 7. In this subset with fractal dimension $D_2 = 2$, the same analysis was performed using the STB-meter and the results are shown in Figure 9. We can verify that the behavior of the subregions of the region **01** is according to the expected fractal dimension as shown in Figure 7.

In order to evaluate the STB-meter method with real data, we conducted an experimental study using data streams composed of climatological time series obtained from a network of weather stations with 24 meteorological stations located in the state of São Paulo, Brazil [1]. Data streams are composed of daily measurements of average temperature and precipitation from 1961 to 1990, and the latitude and longitude of each station.

The method was parameterized to use a three-month sliding window, i.e., $n_c = 3$ and the parameter refresh rate was defined as monthly ($n_i = 30$). In order to compute the fractal dimension of the subsets indexed in the STB-tree, we defined queries reaching until the third level.

The STB-tree indexes data for the recent 3 months of the stream; the fractal dimension is then computed at each month using the counting trees of the leaf nodes and aggregating them until the

[1]Provided by Agritempo - www.agritempo.br

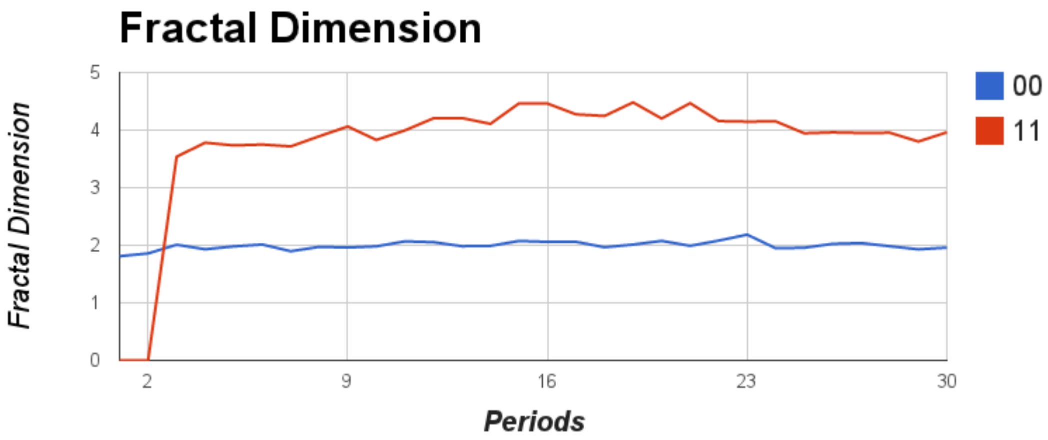

Fig. 9. Fractal dimension of region **00**.

fractal dimension of the highest levels is calculated.

Figure 10 presents a hierarchical data structure (quadtree like) built on the spatial information from 24 meteorological stations in the state of São Paulo.

Thus, results obtained are series of D_2 varying from local behavior of each data subset until the global behavior of the data set. These results may be visualized in Figure 11 where all the graphs of the variation of the fractal dimension for all 3 levels of query on the quadtree are presented.

Figures 12, 13 and 14 show respectively the fractal dimension of the data sets of quadtree for Northwest, Southeast and Southwest regions. The Northeast region was not considered because it contains only one point, having no representation in the general behavior of the data set.

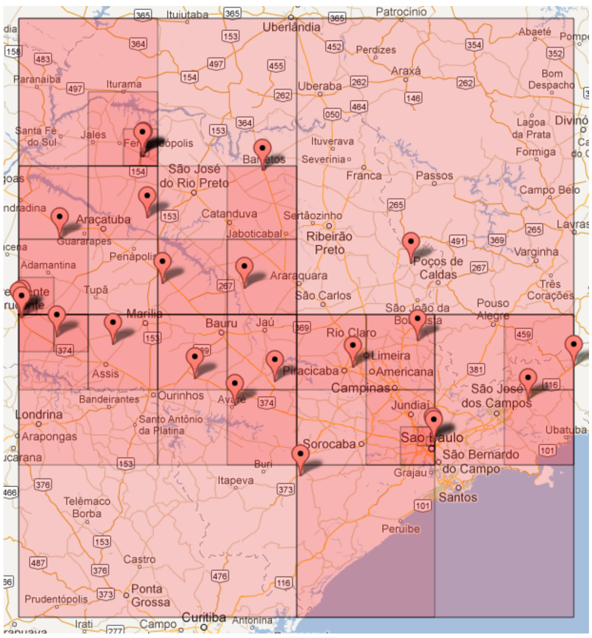

Fig. 10. Quadtree built on the latitudes and longitudes of the 24 meteorological stations of the state of São Paulo.

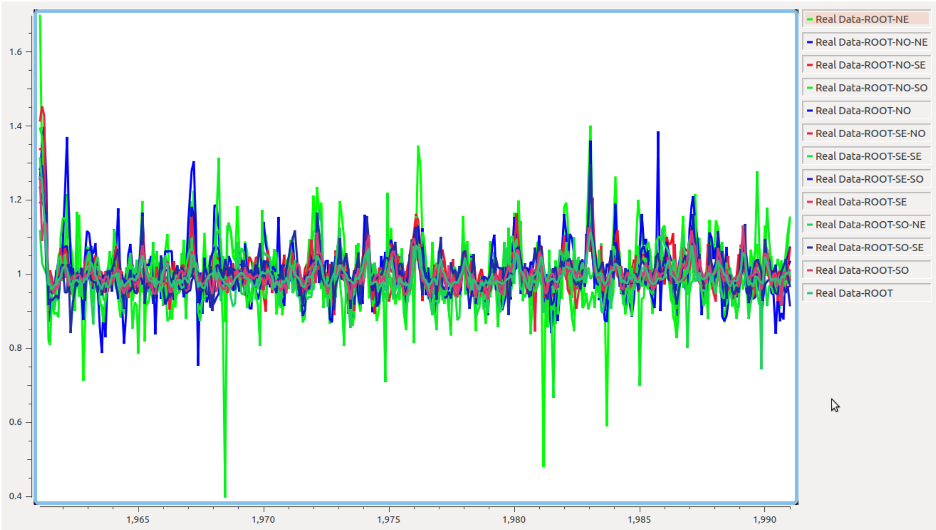

Fig. 11. Fractal dimension for all subsets.

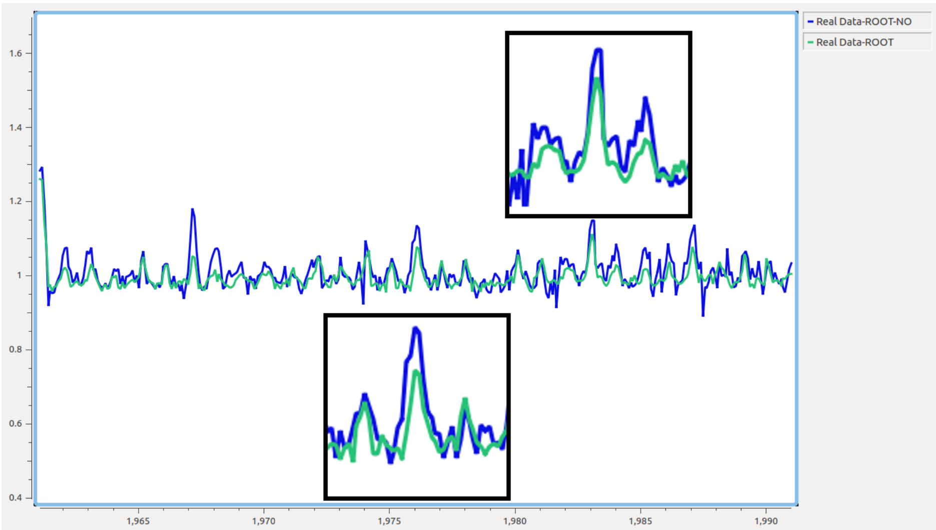

Fig. 12. Both fractal dimension from the Northwest and the state of São Paulo.

The resulting series of fractal dimension suggest that the Southeast is less influenced by extreme events than other regions of the state of São Paulo. The highlighted regions in the figures are periods of low correlation between the variables that compose the data set, i.e., low correlation between precipitation and temperature. According to specialists, those periods are related to extreme climate events resulting from El Niño [Nunes et al. 2011]. It is possible to see that this extreme occurs differently in each subregion.

In these analyses we observed that although precipitation shows more variation than temperature, the algorithm also detected important difference in temperature average, showing that this methodology can detect from small to extreme events of climatic variation. Nowadays, this is relevant, especially in climate change analyses and detection. In addition, this result is also very useful in agrometeorology science. Agricultural production is highly vulnerable to climate variations, and can be considered as the most weather-dependent of all human activities. Small variations in temperature

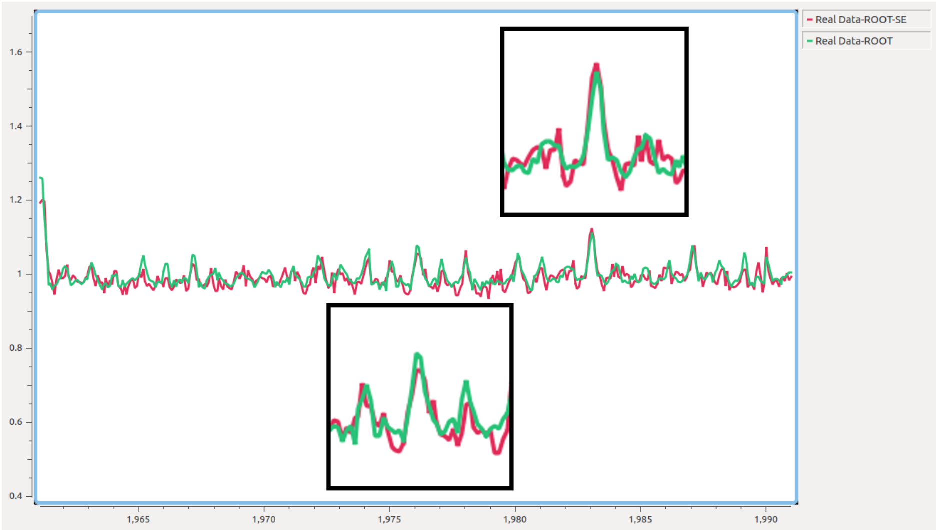

Fig. 13. Both fractal dimension from the Southeast and the state of São Paulo.

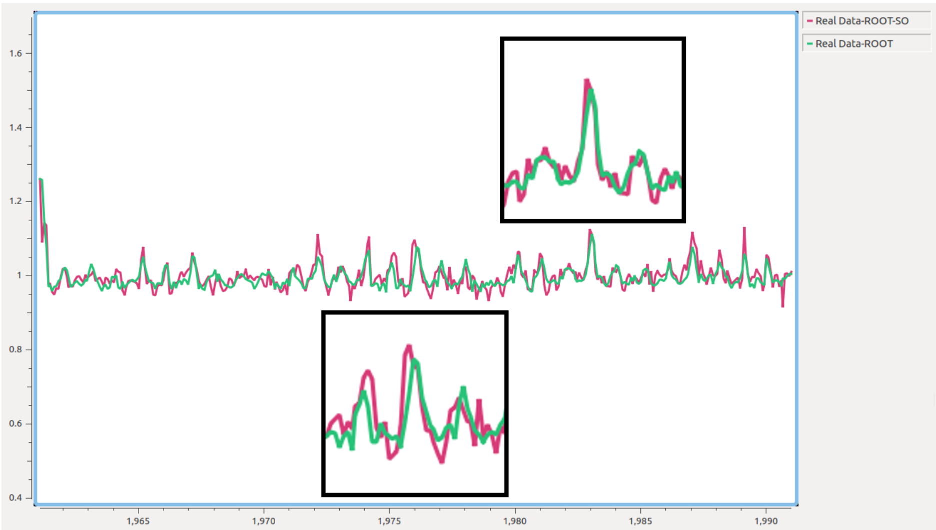

Fig. 14. Both fractal dimension from the Southwest and the state of São Paulo.

(witch normally is difficult to observe in most of analyses) can cause significant losses in yield. In this context, detection of climate variation and extreme climate events is vital to predict and monitor yield.

Analyzing the variation of fractal dimension for different subregions of the data set, it is possible to evaluate the behavior of climate variables over time, for different regions of the state of São Paulo. That analysis allows the specialist to understand how each subregion of the dataset behaves, helping to identify patterns of behavior spatial and temporal climate, as well as understand how each subregion reacts to a specific event.

5. CONCLUSIONS

In this article we present the *STB-meter*, a new method for spatio-temporal analysis of multidimensional data streams. Our approach combines fractal-based analysis to monitor temporal behavior with a multi-resolution, hierarchical structure to deal with spatial properties. Moreover, *STB-meter* handles multidimensional data streams with a single pass on data and feasible computational cost.

Experimental results on synthetic data and real climate data show that our solution is able to spot spatio-temporal behavior patterns, especially anomalies such as extreme climate events. Furthermore, it supports a multi-resolution analysis of temporal and spatial properties as well.

Results show that STB-meter can identify important patterns and extreme events in climate data. This framework is very useful in applied sciences, as agrometeorology and climatology, especially climate change applied to agriculture. Temperature and precipitation are important variables in agricultural analyses, but the huge volume of available data makes the work of specialists very difficult. Once the climate patterns can be identified quickly and easily, specialists have an important tool to monitor and predict yield, in present and future climate conditions.

As future work, we intend to apply the method to other application domains, performing experiments with other data sets. Another relevant point to be explored is to use other hierarchical (spatial) data structures according to the characteristics of the data to be analyzed.

REFERENCES

Baioco, G. B., Traina, A. J. M., and Traina, C. Mamcost: global and local estimates leading to robust cost estimation of similarity queries. In *Proceedings of the International Conference on Scientific and Statistical Database Management*. Banff, Canada, pp. 6–16, 2007.

Barbará, D. and Chen, P. Fractal Mining - self similarity-based clustering and its applications. In *Data Mining and Knowledge Discovery Handbook*. Springer, pp. 573–589, 2010.

Böhm, C. A Cost Model for Query Processing in High Dimensional Data Spaces. *ACM Transactions on Database Systems* 25 (2): 129–178, 2000.

Böhm, C., Berchtold, S., and Keim, D. A. Searching in High-dimensional Spaces-index Structures for Improving the Performance of Multimedia Databases. *ACM Computing Surveys* 33 (3): 322–373, September, 2001.

Chakrabarti, D. and Faloutsos, C. F4: large-scale automated forecasting using fractals. In *International Conference on Information and Knowledge Management*. McLean, EUA, pp. 2–9, 2002.

Cordeiro, R. L. F., Traina, A. J. M., Faloutsos, C., and Traina Jr., C. Halite: fast and scalable multiresolution local-correlation clustering. *IEEE Transactions on Knowledge and Data Engineering.* 25 (2): 387–401, 2013.

de Sousa, E. P. M., Jr., C. T., Traina, A. J. M., Wu, L., and Faloutsos, C. A Fast and Effective Method to Find Correlations among Attributes in Databases. *Data Mining And Knowledge Discovery* 14 (3): 367–407, 2007.

Faloutsos, C. and Kamel, I. Beyond Uniformity and Independence: analysis of r-trees using the concept of fractal dimension. In *Proceedings of the International Symposium on Principles of Database Systems*. Minneapolis, USA, pp. 4–13, 1994.

Faloutsos, C., Seeger, B., Traina, A. J. M., and Traina, C. Spatial Join Selectivity Using Power Laws. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*. Dallas, USA, pp. 177–188, 2000.

Gaede, V. and Günther, O. Multidimensional Access Methods. *ACM Computing Surveys* 30 (2): 170–231, 1998.

Gama, J. *Knowledge Discovery from Data Streams*. Chapman & Hall/CRC, 2010.

Meng, Y. and Dunham, M. Mining Developing Trends of Dynamic Spatiotemporal Data Streams. *Journal of Computers* 1 (3): 43–50, 2006.

Mokbel, M. F. and Aref, W. G. Sole: scalable on-line execution of continuous queries on spatio-temporal data streams. *The International Journal on Very Large Data Bases* 17 (5): 971–995, 2008.

Nunes, S. A., Romani, L. A. S., de Ávila, A. M. H., Jr., C. T., de Sousa, E. P. M., and Traina, A. J. M. Fractal-based Analysis to Identify Trend Changes in Multiple Climate Time Series. *Journal of Information and Data Management* 2 (1): 51–58, 2011.

Samet, H. The Quadtree and Related Hierarchical Data Structures. *ACM Computing Surveys* 16 (2): 187–260, 1984.

Schroeder, M. *Fractals, Chaos, Power Laws: minutes from an infinite paradise*. W. H. Freeman, 1991.

Sousa, E. P. M., Traina, A. J. M., Traina, C., and Faloutsos, C. Measuring Evolving Data Streams' Behavior Through Their Intrinsic Dimension. *New Generation Computing* 25 (1): 33–60, 2007.

Traina, C., Traina, A. J. M., Wu, L., and Faloutsos, C. Fast Feature Selection Using Fractal Dimension. In *Proceedings of the Brazilian Symposium on Database Systems*. CEFET-PB, João Pessoa, Brazil, pp. 158–171, 2000.

Traina Jr., C., Sousa, E. P. M., and Traina, A. J. M. Using Fractals in Data Mining. In M. M. Kantardzic and J. Zurada (Eds.), *Next Generation of Data Mining Applications*. Vol. 1. Wiley/IEEE Press, pp. 599–630, 2005.

Yoon, S. and Shahabi, C. The Clustered AGgregation (CAG) Technique Leveraging Spatial and Temporal Correlations in Wireless Sensor Networks. *ACM Transactions on Sensor Networks* 3 (1): 3:1–3:39, 2007.