



Visual analytics for spatiotemporal events

Ricardo Almeida Silva^{1,2} · João Moura Pires² · Nuno Datia^{1,2}  ·
Maribel Yasmina Santos³ · Bruno Martins⁴ · Fernando Birra²

Received: 21 December 2018 / Revised: 5 July 2019 / Accepted: 17 July 2019 /
Published online: 16 August 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Crimes, forest fires, accidents, infectious diseases, or human interactions with mobile devices (e.g., tweets) are being logged as spatiotemporal events. For each event, its geographic location, time and related attributes are known with high levels of detail (LoDs). The LoD plays a crucial role when analyzing data, as it can highlight useful patterns or insights and enhance the user's perception of phenomena. For this reason, modeling phenomena at different LoDs is needed to increase the analytical value of the data, as there is no exclusive LoD at which the data can be analyzed. Current practices work mainly on a single LoD of the phenomena, driven by the analysts' perception, ignoring that identifying the suitable LoDs is a key issue for pointing relevant patterns. This article presents a Visual Analytics approach called *VAST*, that allows users to simultaneously inspect a phenomenon at different LoDs, helping them to see in what LoDs do interesting patterns emerge, or in what LoDs the perception of the phenomenon is different. In this way, the analysis of vast amounts of spatiotemporal events is assisted, guiding the user in this process. The use of several synthetic and real datasets supported the evaluation and validation of *VAST*, suggesting LoDs with different interesting spatiotemporal patterns and pointing the type of expected patterns.

Keywords Data visualization · Spatiotemporal patterns · Multiple levels of detail · Visual analytics

1 Introduction

A spatiotemporal event is a particular activity that occurred in space and time. For example, *homicide*((41.878037, -87.629442), 09/05/2015 20:00, 2) can represent a homicide that occurred at the latitude and longitude coordinates (41.878037, -87.629442), at eight o'clock

✉ Nuno Datia
datia@isel.pt

¹ ISEL, Instituto Politécnico de Lisboa, Lisbon, Portugal

² NOVA LINCS, FCT, Universidade NOVA de Lisboa, Lisbon, Portugal

³ ALGORITMI Research Centre, University of Minho, Braga, Portugal

⁴ INESC-ID and IST, University of Lisbon, Lisbon, Portugal

PM of a given day, resulting in two victims. Spatiotemporal events can be described as data with the following structure: $event(S, T, A_1, \dots, A_N)$, where S describes the geographic location of the event, T specifies the time instant/interval, and A_1, \dots, A_N are attributes detailing what has happened. Spatiotemporal patterns are non-uniform distributions of events across the space and time. Finding such spatiotemporal patterns helps to understand the associated phenomena [26, 34, 48].

Nowadays, Visual Analytics (VA) approaches targeting the analysis of spatiotemporal events have been developed to analyse a single phenomenon (e.g., crimes), focusing on a specific kind of pattern, such as spatiotemporal hotspots [12, 15, 40]. However, patterns might appear in many different forms [53]. For example, in some phenomena, events form spatiotemporal clusters (e.g., tweets); in others, events form a cloud that moves in space throughout time (e.g., spreading of a disease); or, events occur spread-out throughout space, with some regions revealing a higher intensity (e.g., robberies). Therefore, focusing on a specific type of pattern may leave many others undetected.

We also have that most VA approaches have been designed to follow a single Level of Detail (LoD) analysis [38, 70]. Nevertheless, the LoD matters for the perception of patterns, and often there is no exclusive LoD to study phenomena [23, 29, 49, 58]. Although the LoD plays a crucial role in the perception of patterns, users have been left with the choice of the LoD to look for patterns.

At an early stage of analysis, when users are not familiar with a spatiotemporal phenomenon, users often face difficulties to identify the LoDs in which patterns can be better perceived. They can easily fall into a condition of information overload [29], where users likely face difficulties to identify the LoDs in which patterns can be better perceived.

To enhance analyses over spatiotemporal events, we propose to move from a single user-driven LoD to a multiple LoDs analysis approach. This provides to the user an understandable high-level overview of the phenomenon underlying structure at different LoD, with hints about the distribution of the events in space and/or in time and a glimpse of the presence or absence of patterns. Following this approach, the user might detect very soon in what LoDs there are potential patterns and there kind.

The present article describes a web-based VA tool, named *VAST* (Visual Analytics for SpatioTemporal events), anchored on the *SUITE* framework [55]. *VAST* allows users to have hints about the absence or presence of different kinds of spatiotemporal patterns at multiple LoDs. To the best of our knowledge, there is no other approach that is independent of from the application domain and that simultaneously supports analyses over spatiotemporal events at multiple LoDs.

The evaluation of our proposal was conducted with two types of datasets of spatiotemporal events, namely (i) synthetic datasets, and (ii) real datasets. Synthetic datasets with different spatiotemporal patterns, at different LoDs, were produced. For most cases, *VAST* could provide a correct overview of the phenomenon, with the identification of the LoDs in which patterns exists and, therefore, the LoDs that should be used to detail the analysis. The real datasets studied were: (i) forest fires in Portugal; and (ii) violent attacks against civilians occurring in Africa; *VAST* was effective in identifying patterns for both datasets, at different spatiotemporal LoDs.

The rest of the article is organized as follows. The relevant related work is summarized in Section 2. Section 3 introduces the background information about the granular theory for representing spatiotemporal data at different spatiotemporal LoDs. Section 4 details the interface of *VAST*, describing the way phenomena are analyzed at multiple LoDs. Section 5 presents the experiments carried out using synthetic and real datasets. Finally, we conclude and point out directions for future work in Section 6.

2 Related work

Many approaches have been proposed in the literature that allow analyses over spatiotemporal events, in different research areas.

2.1 Information visualization approaches

To understand the dynamic of spatiotemporal events, animated maps [5] and change maps [5] are often used. However, maps only represent multi-attribute data and dynamism [1, 6]; change maps are limited to small amounts of data and a few snapshots (each map representing a time instant or a time interval); the effectiveness of animated maps is therefore compromised [65].

The role of visualization is an open issue when dealing with numerous spatiotemporal events at high LoDs. The visualization methods get easily cluttered and become difficult to analyze [37, 54]. Visualization methods allowing the understanding of spatiotemporal events at different LoDs are still an open issue that in information visualization's literature. This happens because a visualization needs to combine the spatial and temporal dimensions in a smart way in order to be understandable, a task that is quite challenging.

Aigner et al [1] make a comprehensive survey of techniques used for visualizing time-oriented data¹. From the 115 visualization methods surveyed by [1], just 19 were designed to display spatiotemporal data. From these 19, 4 (Flow Map, Flowstrates, Space-time Path, and Trajectory Wall) were designed to show movements of objects over time, which is out of the scope of this work.

From the remaining 15, 4 (GeoTime [28], Space-time Cube [32], Time Varying-Hierarchies on Maps [25], and Spatio-temporal Event Visualization [21]) make use of the space-time cube concept (X-Y to represent latitude and longitude and Z to represent time). In particular, the Spatio-temporal Event Visualization [21] was designed specifically for displaying spatiotemporal events so that they are placed within the space-time cube and the event's attributes can be encoded with visual variables like size or color, among others [7]. However, 3D visualizations commonly suffer from occlusion and overplotting, making it difficult to grasp spatiotemporal patterns from a visual inspection.

A similar issue emerges from the 4 visualization methods (Data Vases [61], Helix Icons [64], Pencil Icons [64], and Wakame [16]) that use 3D diagrams over geographic regions, as well as from the 2 visualization methods (Icons on Map [17], and Value Flow Map [4]) that use 2D diagrams to map the corresponding data values varying over time. Notice that, in order to use these last 6 visualization methods in the context of spatiotemporal events, we have to aggregate them by geographic regions. However, the diagrams can have a difficult readability if the number of geographic regions under study is high, or if they are quite close to each other.

The Time-oriented Polygons [52] visualization may also have readability problems. This approach creates a partition of each polygon (2D) where each partition maps a value regarding a time period (using the color). The readability problems will emerge when one considers small polygons or/and many time-periods. From the remaining results, the most relevant methods for the analysis of spatiotemporal events might be: the Great Wall of Space-time [63], VIS-STAMP [24], and Growth Ring Maps [2].

¹The website: www.timeviz.net

VIS-STAMP [24] is actually a visual analytics software package that couples computational, visual, and cartographic methods for exploring and understanding spatiotemporal and multivariate data. Although this approach allows the search for spatiotemporal patterns, this can only be done for one spatiotemporal LoD at a time.

The Great Wall of Space-time [63] creates a 3D wall based on a topological path over a cartographic representation. This wall is used to display how the data values of the geographic regions belonging to the path vary over time. This approach is not suitable to analyze spatiotemporal events because they are often spread out in space and time. Therefore, we are not generally interested in a particular spatial path to analyze the phenomenon.

Growth Ring Maps [2] is a technique for visualizing the spatiotemporal distribution of events. Every spatiotemporal event is represented by one pixel. Each location (for example the centroid of spatial clusters of events) is taken as the center point for the computation of growth rings. The pixels (i.e., events) are placed around this center point in an orbital manner resulting in the so called Growth Ring representations. The pixels are sorted by the time at which the event occurred: the earlier an event happened, the closer to the central point the pixel is. Although this approach can be useful to provide a grasp on the spatiotemporal distribution of events, a clear understanding about when spatiotemporal hotspots occurred can be hard to achieve through visual inspection. Furthermore, there might be other patterns that are not captured like changes in the structure of the spatial distribution of events throughout time.

As mentioned, the design of a visualization method that aims to combine the spatial and temporal dimensions of data is not trivial. Perhaps that is why from 115 visualization methods surveyed by [1], we only have 19 visualization methods for spatiotemporal data. Their usage for spatiotemporal events was further discussed in this work and, in short, they have some problems handling spatiotemporal events. Another characteristic which is transversal to the visualization methods discussed here is that they encode data into visual representations at certain LoDs. In fact, from our perspective, the visualizations should be used according to the LoD of the input data, in spite of the issues identified for using them. For instance, Spatiotemporal Event Visualization [21] should be used when spatiotemporal events are provided at high LoDs (latitude and longitude coordinates), while the Time-oriented Polygons [52] approach should be used when the events are aggregated by some administrative level (e.g., counties) and by year.

In general, a visualization method produces a single representation of data. In order to make this representation effective, the visualization methods are designed taking into account the analytical goal and sometimes the data [1]. However, the analysis of spatiotemporal data frequently requires coordinated views in order to deal simultaneously with the spatial, temporal, and thematic aspects of the data [13]. This approach has become standard in the recent visual analysis applications because it directly support the expression of complex queries using simple interactions [13, 51, 67].

2.2 Automated approaches

Shekhar et al. [53] provide a survey about spatiotemporal pattern families. The main families identified are spatiotemporal outliers, spatiotemporal coupling, spatiotemporal partitioning or summarization, and spatiotemporal hotspots.

A spatiotemporal outlier is a spatially and temporally referenced object or event whose non-spatiotemporal attribute values differ significantly from those of other objects in its spatiotemporal neighborhood. For example, spatiotemporal outlier detection can be used

to detect the occurrence of unexpected events like crimes or traffic accidents. Spatiotemporal coupling patterns represent spatiotemporal objects or events which occur in close geographic and temporal proximity. For example, analysis of crime datasets may reveal frequent occurrence of misbehaviors and drunk driving after and near bar closings on weekends. Spatiotemporal clustering is the process of grouping similar spatiotemporal objects or events, partitioning the underlying space and time. For example, partitioning and summarizing crime data, which is spatial and temporal in nature, helps law enforcement agencies to find trends of crimes and effectively deploy their police resources [11, 41]. Spatiotemporal hotspots are neighbouring regions jointly with certain time intervals where the number of objects or events is anomalously or unexpectedly high. For example, in epidemiology, finding disease hotspots allows officials to detect an epidemic and allocate resources to limit its spread [19].

Several algorithms have been developed to compute spatial and spatiotemporal patterns and several survey can be found in the literature [20, 43, 50, 53].

Often, the patterns have statistical expression. This way, spatial or spatiotemporal statistics are proposing quantitative analysis about the presence or absence of such patterns. The average nearest neighbor index [14] (ANN) can give some hints about the presence of spatial clustering. If the ANN value is less than one, the pattern exhibits clustering. Otherwise the trend is toward dispersion. Getis-Ord General G [22] measures how concentrated the high or low values are for a given study area. Positive scores indicate that the spatial distribution of high values is spatially clustered, while negative scores indicate that the spatial distribution of low values is spatially clustered. Getis-Ord General G measure might also suggest spatial outliers [22]. Global Moran's I [45] measures the spatial autocorrelation or dependency based on feature locations and an associated attribute. When the spatial distribution of high values and/or low values in the phenomena is more spatially clustered than would be expected if the underlying spatial processes were random, the Global Moran's I value will be positive. Spatiotemporal statistics like Knox [31], Mantel [42] or the Jacquez k-nearest neighbor test [27] measure the level of spatiotemporal interaction embedded in a phenomenon. More recently, [19] proposed estimators to measure the spatiotemporal clustering/regularity in spatiotemporal point processes (equivalent terminology for spatiotemporal events with points as their spatial representation).

One challenge to mine spatiotemporal data results from the Modifiable Area Unit Problem (MAUP) [47] or the multi-scale (i.e., multiple LoD) effect, since the results depend on a choice of appropriate spatial and temporal scales (i.e., LoDs) [59]. This means that patterns may be biased due to how data is aggregated/summarized. Analyses across multiple LoDs can make the MAUP identifiable or discarded sooner. For example, when a pattern is only visible in a specific LoD it can be further validated. One might conclude that the pattern suffers from MAUP and can be ignored or, if the phenomenon specifically operates there, it can be considered valid. Therefore, we argue that the analysis across multiple LoDs can attenuate the MAUP.

2.3 Visual analytics applications

There are several applications/prototypes to make analyses over spatio-temporal events. In the project carried out by [33], 47 applications / geovisualization methods were assessed. Among the applications studied, 25% (12) were developed to analyze phenomena logged as spatiotemporal events. None of the approaches support data view at multiple spatial and temporal granularities (i.e., spatiotemporal LoDs).

Visual analytic (VA) approaches have also been proposed in the literature. Some of these approaches support separate analyses of space and time anchored on descriptive statistics, most commonly considering one st-LoD at a time.

Lins et al. [38] proposed a compressed hierarchical data structure to hold huge amounts of spatiotemporal events in memory. In addition, the authors implemented some web-based applications to explore real datasets of spatiotemporal events. The spatial LoD at which the events are displayed changes according to the zoom level. However, the same behavior was not registered when the time series was analyzed. Besides that, the interface contains a line chart with the number of events aggregated by day. This approach does not focus on a particular analytical goal, as it uses the descriptive statistic `COUNT`. Another characteristic is the fact that this approach is independent from the phenomenon. Furthermore, and although they have spatiotemporal events available at different spatial and temporal LoDs, their analyses are conducted using one spatial or temporal LoD at a time, separately.

Ferreira et al. [15] developed a visual environment to explore taxi trips, called *TaxiVis*. The input data are events of taxi pickups and taxi drop-offs that happened in New York City. This approach supports exploratory analysis about taxi pickups and taxi drop-offs without any particular analytical task in mind. They are addressed using descriptive statistics that result from the separate analysis of the spatial and the temporal dimension of data.

Some of the VA approaches discussed in the literature support separate analyses of space and time, but these analyses are performed at one spatiotemporal LoD at a time [12, 15, 38]. More approaches were found [3, 12, 30, 39, 41].

Other approaches support analyses that look for spatiotemporal patterns [24] or [40]. However, these kinds of approaches follow analyses based on a single LoD and, in some cases, they are developed for the detection and exploration of a particular spatiotemporal pattern in a particular application domain [10, 60, 62, 66].

As opposed to that, we aim to give an overview of the presence or absence of spatiotemporal patterns at different LoDs simultaneously, without focusing in a particular application domain and by just considering phenomena logged as spatiotemporal events.

2.4 Manifold LoDs approaches

The scale (or LoD) of analysis can highly affect results [46]. This issue has been acknowledged a long time ago [47]. However, with spatiotemporal events in mind, analytical approaches have been mainly developed to support analyses based on a single LoD. Thus, the MAUP becomes an issue, once unsuitable LoDs can hide patterns and conceal the true underlying nature of a dataset.

VA approaches working across LoDs are still in their infancy despite the fact that they have been gaining more attention in recent years [35, 36, 69]. [58] propose a Visual Analytics approach called *Pinus*, aiming at the detection of patterns at multiple temporal LoDs in numerical time series, specifically from the environmental sciences. To accomplish that, statistical measures are computed for all possible time LoDs (i.e., scales) and starting positions, namely, mean, variance, and discrete entropy. This approach makes no assumption about the temporal LoD and the temporal patterns. It combines statistical measures and the pattern recognition abilities of the user to support effective detection of temporal patterns at different temporal LoDs. We aim to bring this mindset for the analysis of spatiotemporal events at several spatiotemporal LoDs.

Goodwin et al. [23] proposed a framework for analyzing multiple variables across spatial LoDs and geographical locations. Based on it, the authors developed a suite of novel interactive visualization methods to identify interdependencies in multivariate data coupled

with a series of correlation matrix views. This approach does not focus on a particular phenomenon and was devised to look for correlations on multiple variables in multiple spatial LoDs and geographic regions.

Robinson et al. [49] developed a visual analytics approach, called STempo, to support the discovery of patterns found in spatiotemporal events. STempo was designed to detect and analyze significant co-occurrences of real-world events. This approach is making a separate analysis of the temporal and spatial dimension of events, as the input for the T-pattern algorithm corresponds to records containing the timestamp and a set of event types that occurred in it. Finally, this approach looks for temporal patterns and not for spatiotemporal patterns, because the sequences identified are not assigned to specific geographic regions. Nevertheless, this approach computes temporal patterns in multiple temporal LoDs.

The visual analytics approaches discussed so far explore time following a linear model. However, periodicity underlies many phenomena. Swedberg and Peuquet [59] proposed a visual analytics web application developed to help users in the detection and analysis of calendar related periodicity in spatiotemporal event data sets via exploratory user interaction. This work allows for the analysis at multiple spatial LoDs and temporal LoDs despite the fact that the number of the spatial LoDs that we can analyze, simultaneously, is limited to two (raw data and aggregated data by the user-defined geographic regions). Although the mentioned patterns are interesting, they are obtained by working with space and time separately using only the COUNT descriptive statistic.

To the best of our knowledge, there are no approaches that work across several spatial and temporal LoDs, working with space and time together and, therefore, looking for spatiotemporal patterns at different spatiotemporal LoDs. Furthermore, the VA approaches discussed here do not have any theoretical foundation that anchors the analysis across LoDs. The approaches rely on clever visual designs that show data at different LoDs. However, from our perspective, a theoretical foundation that anchors the analysis across LoDs can be important for having phenomena representations at different LoDs and, then use better suited visualization methods to display them.

3 Primer on granular theory

Granular computing has emerged as a paradigm of knowledge representation [68], where granules are basic ingredients of information. Roughly, a granularity defines a division of a domain in a set of granules disjoint from each other [56]. *Counties* and *States* are common examples of spatial granularities; *Hours* and *Days* are examples of temporal granularities.

Under a general theory of granularities [56], a granular computing approach was devised to model spatiotemporal phenomena at multiple LoDs. This approach was labeled as the granularities-based model [57], where a phenomenon is modeled through a collection of statements. Granules are used in the statements' arguments. For example, we can model a crime event through the statement: *crime(Oakland, 03/01/2015 18h, 1, homicide)* where the granules used in the statement come from the granularities *County*, *Hour*, *Natural Number*, and *Crime Types*.

Statements are made at some LoD. The set of granularities involved in the statement defines the LoD at which an event is described. For example, the LoD of *crime(Oakland, 03/01/2015 18h, 1, homicide)* is defined by the corresponding granularities: *County*, *Hour*, *Natural Number*, and *Crime Types*. Through the granularities-based model, statements can be generalized to coarser LoDs automatically. Using the granularities-based model, we are able to have a phenomenon modeled in multiple LoDs [57].

Let us consider a statement describing an event using a spatial granule $s \in S$ and a temporal granule $t \in T$. The pair (s, t) is called a spatiotemporal granule (st-granule) of the spatiotemporal LoD (S, T) .

Each st-granule (s, t) indexes the set of statements spatially located at s and temporally located at t . Typically, at a very detailed spatiotemporal LoD (from this point forward referred as st-LoD), events are sparse and mostly non-co-occurring. This means that either the st-granules have no events, or they have just one event. At coarser st-LoDs, the co-occurrence of events on the same st-granule becomes more likely.

On top of the granularities-based model, Silva et al. [55] developed a SUMmarIZIng spatioTemporal Events framework (SUITE) that builds, for each st-LoD, summaries about phenomena represented as spatiotemporal events, called *abstracts*. Abstract values can be a number, a vector, or a matrix. SUITE considers five types of abstracts: (i) Global; (ii) Spatial; (iii) Temporal; (iv) Compact Temporal; and (v) Compact Spatial.

A Global Abstract summarizes all statements by a single abstract value. Known spatiotemporal statistics [19, 53] (e.g., Knox or Mantel statistics) can be used to compute Global Abstracts. A Spatial Abstract summarizes, for each $t \in T$, all statements at t by a single abstract value, so we get a time series of abstract values, each one summarizing the spatial distribution of the events at granule t . Known spatial statistics [14, 53] (e.g., Average Nearest Neighbor or Moran's I) can be used to compute spatial abstracts.

A Temporal Abstract summarizes, for each $s \in S$, all statements at s , by a single abstract value, so we get a map of abstract values, each one summarizing the temporal distribution of the events at granule s . Known temporal statistics [8] can also be used to compute temporal abstracts.

A Compact Temporal Abstract is just a summarization of a Temporal Abstract, i.e., a summarization of a time series of abstract values into a single abstract value. Similarly, a Compact Spatial Abstract is just a summarization of a Spatial Abstract, i.e., a summarization of a map of abstract values into a single abstract value. ST-Abstracts will refer either the Global Abstracts, the Compact Temporal Abstracts, or the Compact Spatial Abstracts.

4 Visual analytics for spatio temporal events

Visual Analytics for SpatioTemporal events (VAST)² was developed to support analysts in the task of visually inspecting the computed abstracts at many LoDs simultaneously, allowing users to understand not only the absence or presence of different kinds of spatiotemporal patterns, but also at which LoDs they are visible or at least in what LoDs they are more likely to be found. VAST implements the granularities-based model and the SUITE framework, including the abstracts presented in Appendix. It is a client-server web-based application coded in Java, providing a set of RESTful Web services (using the Spring framework). It relies on the PostgreSQL with PostGis spatial extension. The browser-based client is coded in JavaScript, HTML5, and uses a WebGL based api [9] to display efficiently thematic maps. VAST is composed by 3 modules: (i) granularities-based module; (ii) SUITE module; (iii) Interface module. The granularities-based and SUITE modules are placed on the server-side, while the interface module is placed on the client-side. Furthermore, the granularities-based and SUITE modules are decoupled from the RESTful Web Service. The application server provides a set of services that are implemented using the interfaces

²VAST prototype can be tested online at <http://staresearch.net/resource#prototypes>.

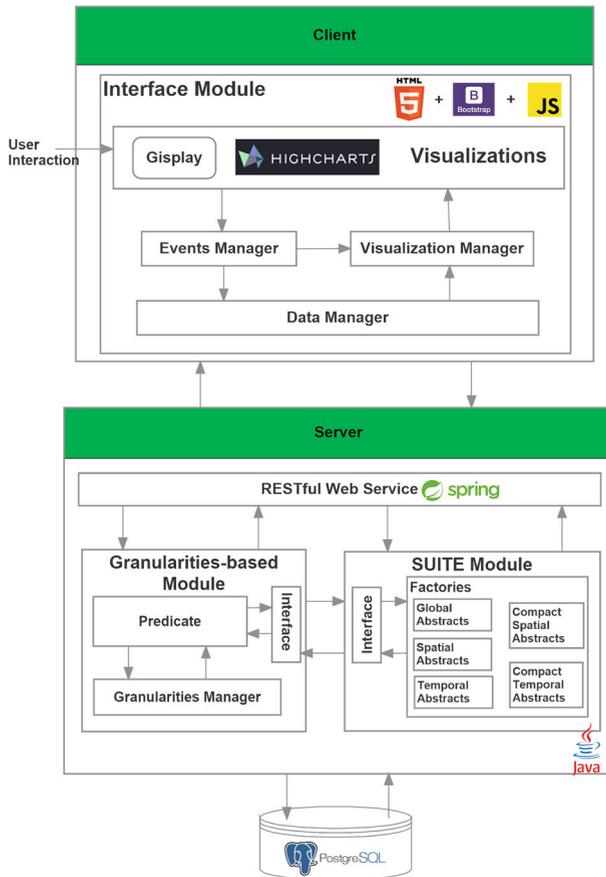


Fig. 1 VAST architecture

exposed by the granularities-based and the SUITE modules. These will be later used by the interface module. An overview of the VAST architecture is illustrated in Fig. 1.

VAST’s design follows the VA Mantra: “Analyze first, show the important, zoom, filter and analyze further, details on demand” [29]. First of all, the interface starts by displaying ST-Abstracts at all available st-LoDs. This interactive visualization may provide hints about different patterns within the spatiotemporal events. Then, one can analyze further by looking at Spatial Abstracts (time series of abstract values) or Temporal Abstracts (i.e., “maps” of abstract values). At any moment of the analysis, it is possible to visually inspect the actual spatial distribution of the phenomenon at a specific temporal granule t in a particular st-LoD.

The interface is composed of three main areas, as illustrated in Fig. 2:

1. **ST-Abstracts;**
2. **Dynamic Abstract Area;**
3. **Phenomena Representation.**

The first area, **ST-Abstracts** (Fig. 2-1), displays a matrix plot for each ST-Abstract. The symbol ▲ points out a Compact Spatial Abstract (e.g., Fig. 2-1.a) while the symbol

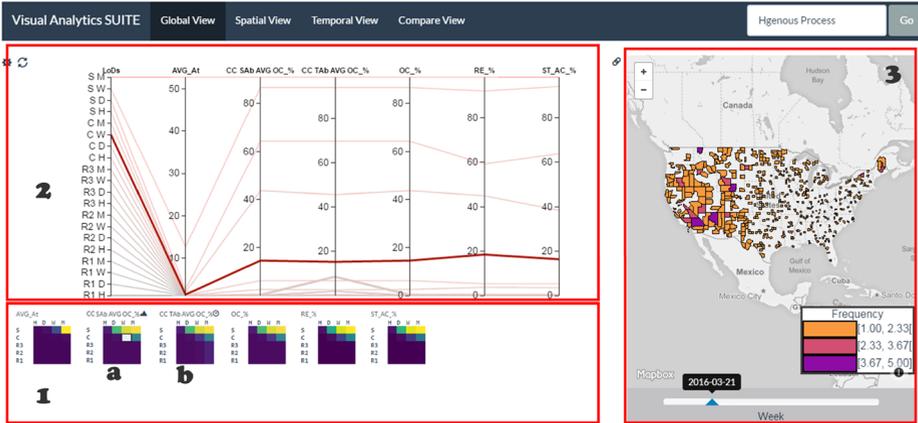


Fig. 2 An overview of the VAST interface

⊙ indicates a Compact Temporal Abstract (e.g., Fig. 2-1.b). When none of these icons is present we have a Global Abstract. Each cell of a matrix plot shows the value of an ST-Abstract at a st-LoD. The skeleton of a matrix plot is displayed in Fig. 3. In the rows, we have the spatial granularities (finer granularities at bottom), and in the columns we have the temporal granularities (finer granularities at left). All used abstract values are numbers and their value is mapped to a color using the color scheme shown in Fig. 3. For example, Fig. 2-1 shows 6 matrices, from left to right:

- a Global Abstract named *Average Atoms in st-granules*;
- a Compact Spatial Abstract named *Average of Spatial Occupation Rate*;
- a Compact Temporal Abstract named *Average of Temporal Occupation Rate*;
- three more Global Abstracts respectively *Occupation rate*, *Reduction rate* and *Collision Rate*.

All matrices are using 5 spatial granularities (State, County, and 3 rasters at different resolutions) and 4 temporal granularities (hour, day, week, and month).

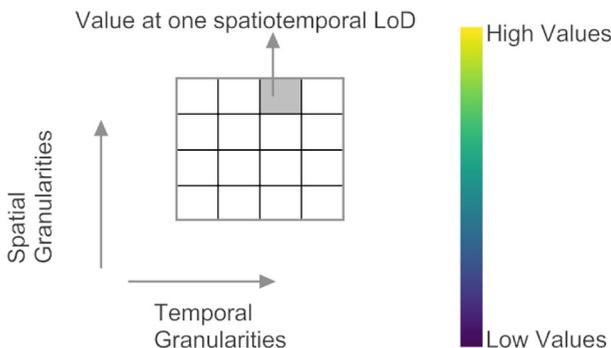


Fig. 3 An overview of the structure of a matrix plot

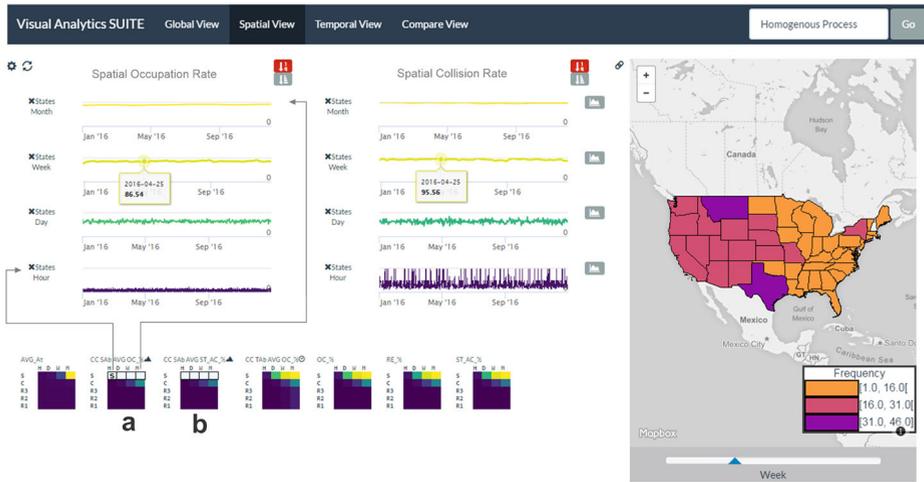


Fig. 4 An overview of the VAST interface with Spatial Abstracts

The **Dynamic Abstract Area** (Fig. 2-2) is used to present 3 different visualizations: (i) a **Global View** that shows a Parallel Coordinates visualization with the same abstracts presented in the ST-Abstracts area; (ii) a **Spatial View** that shows the time series corresponding to a few selected Spatial Abstracts, as illustrated in Fig. 4; and, (iii) a **Temporal View** that shows the maps corresponding to a few selected Temporal Abstracts, as illustrated in Fig. 5.

In the Parallel Coordinates visualization (Fig. 2-2) each line corresponds to one st-LoD. The most left coordinate represents the st-LoDs ordered from the more detailed st-LoD (*R1, Hour*) to the coarser st-LoD (*State, Month*). The other coordinates correspond to the ST-Abstracts presented in ST-Abstract area.

In Fig. 4, there are four cells selected from *Average of Spatial Occupation Rate* (Fig. 4a) and from *Average of Spatial Collision Rate* (Fig. 4b). Therefore, eight Spatial Abstracts are visible in the Spatial View, which are organized/grouped by Spatial Abstract and ordered from the more detailed st-LoD (bottom) to the coarser one (top).

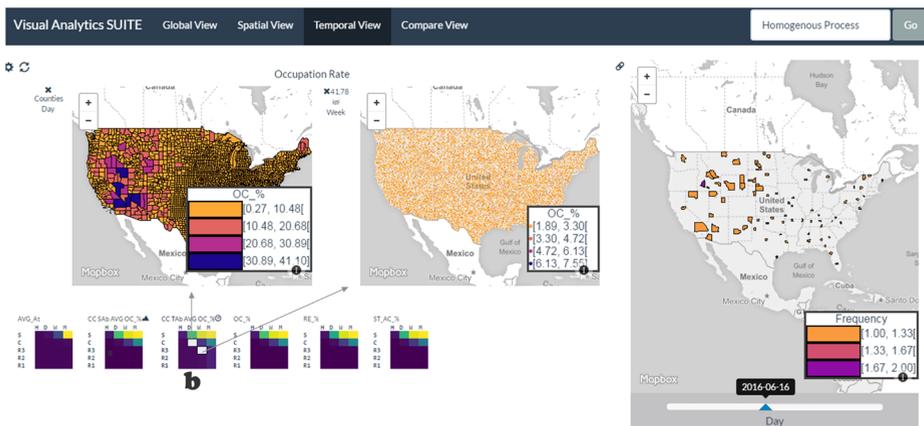


Fig. 5 An overview of the VAST interface with Temporal Abstracts

The Temporal View is illustrated in Fig. 5 and there are two st-LoDs selected from the *Average of Temporal Occupation Rate* (see Fig. 5b): (*R3, Weeks*) and (*County, Day*). As a result, two Temporal Abstracts are displayed. When the st-LoD has a *raster* granularity, the map represents each spatial granule through a point, leading to a dot map (e.g., the map on the right side). Otherwise, the spatial granules are displayed in their original form, which leads to a choropleth map (e.g., the map on the left side).

The last area is the **Phenomena Representation** (Fig. 2-3) used to display spatiotemporal events at a st-LoD using thematic maps. The slider underneath allows the user to scroll temporally through the temporal granules, according to the st-LoD that was chosen. The map displays the number of events for each st-granule.

4.1 Main abstracts implemented

Several abstracts were implemented and actually proposed in the context of this work. A subset of the abstracts implemented/proposed is now described. Whenever some abstract is based on another work, a reference will be placed:

1. The **occupation rate** measures the percentage of spatiotemporal granules occupied, that is, it measures the average density of a model at a given LoD. The value 0 means no spatiotemporal granules is occupied and 100 means that all spatiotemporal granules are occupied.
2. The **collision rate** measures the percentage of the spatiotemporal granules occupied that index more than one event, that is, it measures the average co-occurrence of an event at a given LoD. In this case, 0 means no co-occurrence of events in spatiotemporal granules and 100 means that any spatiotemporal granule has events co-occurring.
3. The **Granular Mantel Bounded and Normalized (GMBN)** score measures the spatiotemporal interaction among granular syntheses. The purpose of this measure is to have a hint of the presence or absence of spatiotemporal clustering patterns or any other pattern that involves spatiotemporal interaction, like a contagious process. The value ranges between 0 and 1, where 0 means no interaction at all among the granular syntheses and 1 means that all the spatiotemporal granules are interacting. The GMBN receives as input parameters the spatial and temporal distances. These distances are expressed in terms of granular extents with respect to the spatiotemporal LoD in which the GMBN is computed.

Spatial Abstracts hold a summary for each temporal granule about the granular syntheses occurred on it. The Spatial Abstracts considered in *VAST* are as follows:

1. The **spatial occupation rate** is computed in the scope of each temporal granule. The values' interpretation is similar to the one presented considering the global abstract. This way, we can track the temporal evolution of the occupation rate.
2. The **spatial frequency rate** measures for each temporal granule the percentage of atoms occurred on it, given all the atoms of the phenomenon at a given LoD. In other words, corresponds to a frequency distribution normalized by the total number of atoms in the phenomenon at a particular LoD. The range of values for this abstract lies between 0 and 1 (in each temporal granule) so that 0 means that no atom occurred on that temporal granule, while 1 means all the atoms occurred on that temporal granule. Through this abstract, we aim to understand how the intensity of the phenomenon

spreads out throughout time. This abstract, developed during the research, is not a novel contribution.

3. The **spatial average nearest neighbor (Spatial ANN)** measures how occupied spatiotemporal granules are dispersed or clustered in each temporal granule. This might indicate variations between dispersed and clustered spatial distributions. The computed value is not a distance but a normalized value such that if the value is less than 1, the spatial pattern might be clustering while if the value is greater than 1, the trend is toward dispersion. Notice that, the z-score³ of the Spatial ANN is also computed. Very low or very high z-score values suggest some spatial pattern and, therefore, we can reject the complete spatial randomness. This abstract was developed based on [14].
4. The **spatial scope** measures the percentage of spatial area occupied by the phenomenon in each temporal granule, where the spatial area is a concave region that encloses all the granular syntheses, and the total spatial area corresponds to the extent of the spatial granularity. Through this abstract, we aim to understand if the spatial scope of the phenomenon varies throughout time.
5. The next Spatial Abstract, named **Spatial Consecutive Distance Between Centers of Mass**, considers two consecutive temporal granules t_{i-1} , t_i . For each one, a region that encloses all spatiotemporal granules is computed. Then, the centroids of each region are computed, and the value of the Spatial Abstract at t_i consists of the distance between the centroid at t_{i-1} and the centroid at t_i . This is done for all temporal granules apart from t_0 where the Spatial Abstract takes the value 0. Through this abstract, we aim to understand whether the phenomenon moves in space throughout time.

Temporal Abstracts hold a summary for each spatial granule about the granular syntheses, considering all temporal scopes. The Temporal Abstracts considered in *VAST* are as follows:

1. The **temporal occupation rate** is, in this case, computed in the scope of each spatial granule. The values' interpretation is similar to the one presented considering the occupation rate. This way, we can assess the occupation rate over the space.
2. The **temporal frequency rate** measures for each spatial granule the percentage of atoms occurred on it, given all events of the phenomenon at a given LoD. The range of values for this abstract lies between 0 and 1 (in each spatial granule) so that 0 means that no event occurred on that spatial granule while 1 means all the events occurred on that spatial granule. This way, we can observe the intensity of the phenomenon over space.
3. A Temporal Abstract, named **temporal average nearest neighbor**, measures how occupied spatiotemporal granules are dispersed or clustered in time, for each spatial granule. The interpretation of values is similar to the one presented in the case of the spatial average nearest neighbor. Furthermore, the corresponding z-score was also implemented.

5 Experiments

VAST was used to conduct experiments over two types of datasets of spatiotemporal events: (i) seven synthetic datasets; (ii) two real datasets.

³<http://mathworld.wolfram.com/z-Score.html>

5.1 Synthetic datasets

The synthetic datasets of spatiotemporal events were generated using the `stpp` R package ([19]). This package exposes a set of functions to simulate spatiotemporal events following different models ([18, 19, 44]):

1. **Homogeneous Poisson Process:** the homogeneous Poisson process is the simplest mechanism for the simulation of a spatiotemporal point pattern. This model hardly approaches a pattern in a phenomenon but provides a good basis for comparison, as it reflects complete spatiotemporal randomness. Informally, in a homogeneous Poisson process, the events form an independent random sample from the uniform distribution on the spatiotemporal domain in which the events were simulated.
2. **Poisson Cluster Process:** the Poisson cluster process simulates spatiotemporal clusters of events. This model might reflect phenomena such as forest fires where several wildfire occurrences appear close in time and space, or the presence of spatiotemporal hotspots of crimes, for instance. Informally, a set of parents are generated and, afterwards, a set of events is generated around each simulated parent. The dispersion of events in space and in time around each parent event is an input parameter through which we specify the st-LoD. In this process, when events happen, they occur near to each other in space and time. However, it is possible that no events occur.
3. **Contagious Process:** A contagious process can be pictured out as a cloud of events moving in space throughout time. The contagion process of a disease, for example, in which the disease is transmitted to other people through direct contact with an infected person. Informally, an initial event is generated and, afterwards, the next generated events are near to locations of the previous event(s) simulated. The spatial and temporal neighborhoods on which the next events are generated are input parameters through which we specify the st-LoD.
4. **Log-Gaussian Cox Process:** The Log-Gaussian Cox process simulates spatiotemporal events such that some regions reveal higher intensity. This model might reflect phenomena that contain geographic regions of higher risk, which might change slowly over time. This pattern might happen with wildfires or infectious diseases, among others. Informally, the Log-Gaussian Cox process is a in-homogeneous Poisson process with a stochastic (i.e., randomly determined) intensity. In this case, we have no precise control of the st-LoD in which the pattern is simulated.

Different datasets were produced following one or more of the models presented above. The set of generated datasets is displayed in Table 1, along with characteristics like the model used to generate the observations, the number of events, and the spatiotemporal LoD in which the pattern/model was simulated. All the datasets were generated within the region of the USA and during one year.

5.1.1 Poisson cluster process

Let us start by Dataset 2. This dataset was simulated with the Poisson Cluster process and is composed by 30.000 events within the region of the USA that occurred during one year. The clusters of events are built around a parent within a spatial distance of 110 km and a temporal distance of one day. For this dataset, the most detailed spatial granularity *Raster* (0.16km^2) is based on a grid of 16384×16384 cells that covers the analyzed spatial extent of the phenomenon, and each cell has an area of 0.16 km^2 . The coarser spatial granularities were obtained by dividing by a factor of 4 the number of cells in the grid. The valid granularities

Table 1 Datasets of spatiotemporal simulated events

	Model	Number Events	LoD
Dataset 1	Homogenous	30.000	NA
Dataset 2	Poisson Cluster	30.000	110 Km, Day
Dataset 3	Poisson Cluster	30.000	2 Km, Week
Dataset 4	Poisson Cluster + Homogenous	33.000	110 Km, Day
Dataset 5	Poisson Cluster	30.000	570 Km, Week
Dataset 6	Contagious	5.000	110 Km, Week
Dataset 7	Log-Gaussian Cox	15.000	NA

for space were thus rasters with cell sizes approximately of 0.16 km^2 (R1), 2.55 km^2 (R2), and 40.74 km^2 (R3). The granularities *County* and *State* were also included. The time granularities used were *Hour*, *Day*, *Week*, *Month*.

The raw data (events) were encoded at the finer st-LoD, which includes the time granularity *Hour* and the space granularity R1. After that, the granularities-based module was used to automatically produce the data for all LoDs, and the *VAST* was used to precompute all the defined abstracts for each LoD.

Figure 6 shows the global abstracts (i.e., the *Occupation rate*, the *Collision rate* and the *GMBN*) for all the st-LoDs of Dataset 2.

The *GMBN* points to the st-LoD (R3, *Day*) as the one with greatest spatiotemporal interaction. This seems to be compliant with the st-LoD in which the pattern was simulated. Regarding the other global abstracts (i.e., the *Occupation rate*, and the *Collision rate*), their values increase when moving towards coarser st-LoDs. This happens because when moving to coarser st-LoD, the co-occurrence of granular syntheses in spatiotemporal granules increases since the number of spatiotemporal granules available at coarser st-LoD decreases. Nevertheless, according to the phenomenon, the values of *Occupation rate* and *Collision rate* might increase at different rates.

To better understand in what st-LoD the perception of the phenomenon distinguishes itself, we use an instrument from the interface module that allows us to correlate two global abstracts.

We have implemented two forms of observing the correlation between two global abstracts. One of them is named correlation evolution through spatial granularities, which allows to observe for each spatial granularity how the correlation behaves, considering all

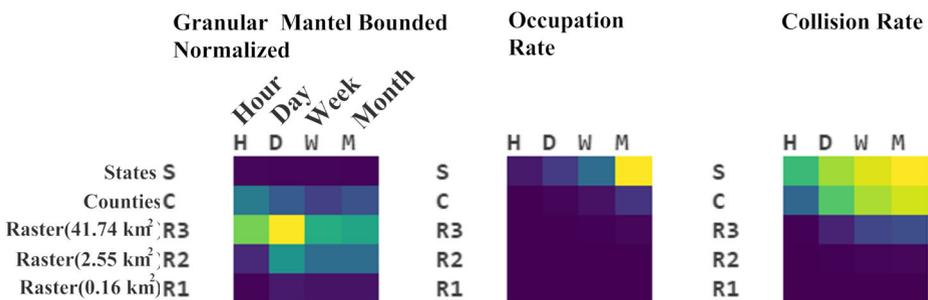


Fig. 6 Global abstracts: *GMBN*, *Occupation rate* and *Collision rate* describing dataset 2

the temporal granularities. The other is named correlation evolution through temporal granularities, which allows us to observe for each temporal granularity how the correlation behaves with respect to all the spatial granularities.

Figure 7a illustrates the correlation evolution through spatial granularities between the GMBN and the collision rate. Each spatial granularity corresponds to a line in the chart. On the other hand, Fig. 7b illustrates the correlation evolution through temporal granularities between the GMBN and the collision rate. In this case, each temporal granularity corresponds to a set in the chart. The color encodes the spatial granularity while the shape of the markers encodes the temporal granularity. This encoding scheme is the same on both forms of correlation. Therefore, a marker with a particular color and shape represents the same spatiotemporal LoD on both charts.

Moreover, in the correlation evolution through spatial granularities the lines connect markers with the same color (i.e., the spatial granularity is the same) while in the correlation evolution through temporal granularities the lines connect markers with the same shape (i.e., the temporal granularity is the same). Notice that both charts might become cluttered according to the mapped data. To attenuate that problem, a user can hide or make visible series of the chart, interacting with the corresponding legend.

On both charts we can observe “elbows”. An “elbow” tip, in these charts, has a particularity that it might be interesting to explore. For the discussion that follows, let us assume that an elbow is created by going from a finer granularity to a coarser granularity (e.g., as happens in the series regarding the spatial granularity $40,74\text{km}^2$ (R3) in Fig. 7a). In these cases, it seems that there is a granularity G such that: (i) for granularities finer than G the correlation seems to be positive; (ii) for granularities coarser than G the correlation seems to be negative. This might be a hint about the LoDs in which the perception of a phenomenon distinguishes itself, considering the two global abstracts at study.

In Fig. 7a, an “elbow” is visible taking into account the spatial granularity R3, where the elbow tip is reached at the granularity *Day*. In Fig. 7b, the most pronounced “elbow” is

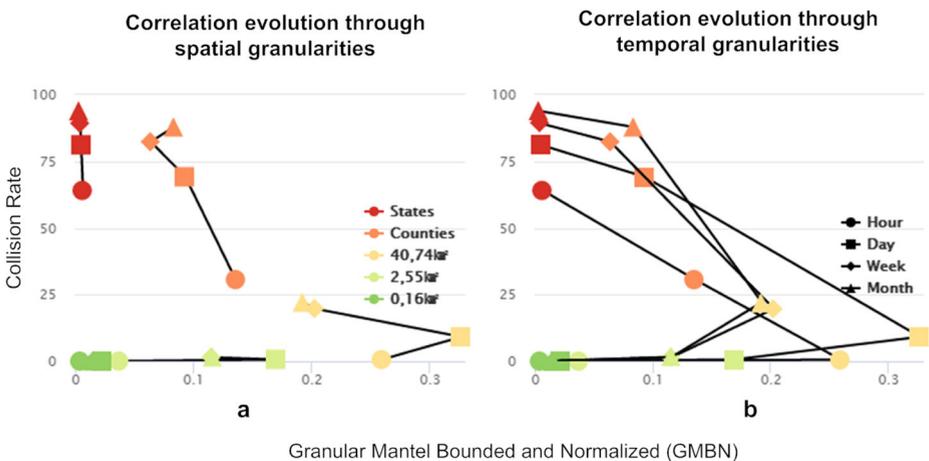


Fig. 7 Correlation between the GMBN and the collision rate

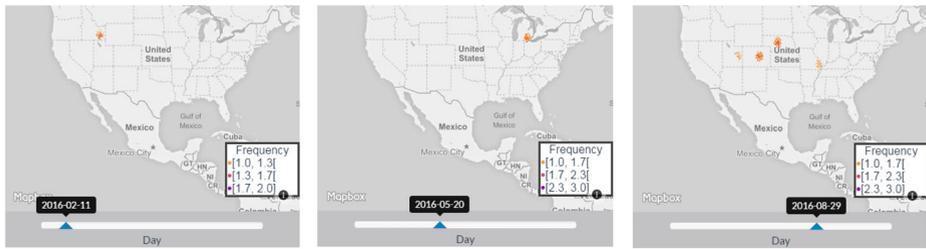


Fig. 8 Dataset 2 at the spatiotemporal LoD R3 and Day, displayed in 3 temporal granules

revealed at the temporal granularity *Day*, where the “elbow” tip is reached at the granularity R3.

Therefore, the st-LoD (R3, Day) is where the “elbow” tip is observed on both charts. This conclusion is similar to the one achieved by just looking at the GMBN, in Fig. 6, and this analysis might seem useless. However, looking only at one Global Abstract as a way of understanding suitable st-LoDs to detail our analyses might be misleading. These scenarios will be discussed later.

The correlation between the GMBN and the collision rate serves two purposes. First, there is one more hint pointing to (R3, Day) as a suitable st-LoD to analyze the data. Second, it allows us to introduce the correlation charts.

Given the evidences pointing that there might be a pattern in the st-LoD - (R3, Day), or at least the phenomenon is observable in such st-LoD, we use the *Phenomenon Representation* area to have a grasp of the data at such st-LoD. The data at three different temporal granules, chosen without any particular criterion, is displayed in Fig. 8. As can be seen, there are clusters of events happening over the USA.

The analysis made so far points out that Dataset 2 (see Table 1) might have a spatiotemporal pattern, and this pattern might be better perceived at R3. The pattern in question are clusters of events happening over time. Our analyses were further detailed using the Spatial and the Temporal Abstracts to confirm a pattern in the st-LoD (R3, Day).

We start by looking to the Temporal Abstract - Temporal Center Mass’s Positioning for three st-LoD as can be seen in Fig. 9. The st-LoD are: (R3, Day), (Counties, Day) and (States, Day). Orange means that most of the events that occurred in the spatial granule are old, while dark blue means that most of the events occurred in the spatial granule are recent in what concerns the extent of the temporal granularity.

Looking at the st-LoD (R3, Day) and (County, Day), in Fig. 9, the geographic regions where the clusters of events have happened can be identified, since spatial granules close to each other have similar values of the temporal center mass’s positioning. In other words, the events occurring near in space seems to occur near in time.

The previous conclusions are also captured by the two Compact Temporal Abstracts of the temporal center mass’s positioning, i.e., its coefficient of variation and its spatial autocorrelation. In this case, the coefficient of variation tells us in what st-LoD the value of the Temporal Center Mass’s Positioning varies more among the spatial granules, while the spatial autocorrelation measures how the value of the Temporal Center Mass’s Positioning is similar in neighboring spatial granules. Thus, we are interested in st-LoDs where there is a considerable variation, and where the spatial autocorrelation’s value suggests spatial

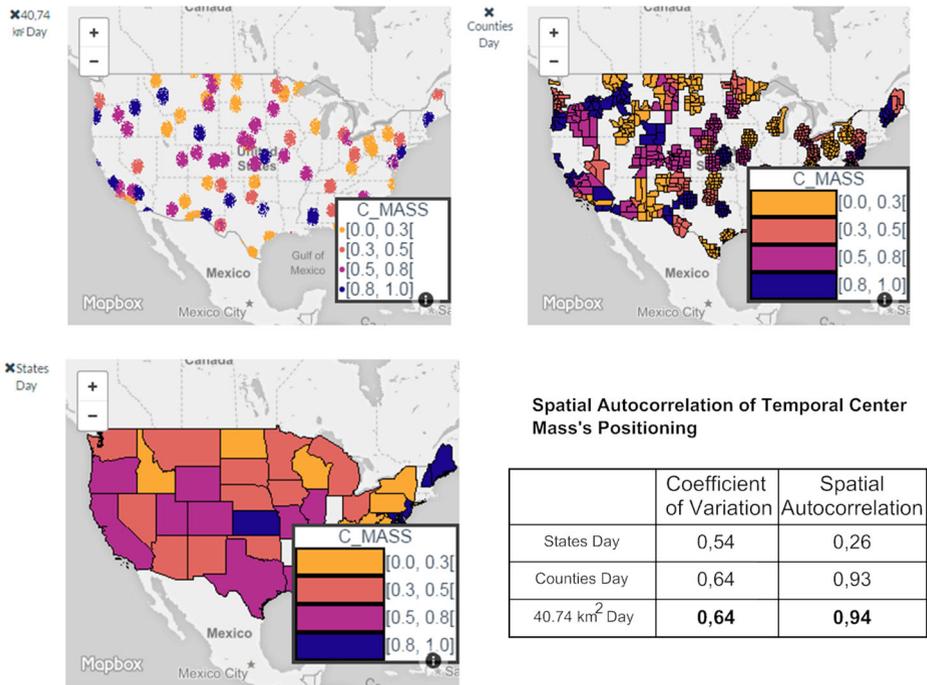


Fig. 9 The temporal center mass's positioning for three st-LoD

correlation. In what concerns the three st-LoD displayed in Fig. 9, the st-LoD ($R3, Day$) is where the Coefficient of Variation and the Spatial autocorrelation take the highest values as detailed in Fig. 9. The spatial autocorrelation is 0.94 (strong positive correlation) and the coefficient of variation is 0.64. Clusters are spread out across the entire country. Besides that, we can relate the geographic regions and the time moments in which the clusters occurred. This kind of perception is lost looking at the data in the st-LoD ($State, Day$) (see Fig. 9), for example.

Since clusters are happening over time, we use the Compact Spatial Abstract - spatial average nearest neighbor (Spatial ANN) - and the corresponding z-score to understand when those clusters of events are happening.

Four st-LoD were chosen: ($R3, Hour$), ($R3, Day$), ($R3, Week$), ($R3, Month$). These were chosen because, based on evidence, we know that st-LoD ($R3, Day$) is appropriate to analyze the data. So, the st-LoD ($R3, Day$) is included in the next analysis. This leaves us with the possibility of varying the spatial or the temporal granularity. Moreover, the previous analysis pointed that the spatial granularity $R3$ was able to show the places where the clusters happened. For this reason, we vary the temporal granularity.

The Spatial Abstracts are displayed in Fig. 10. Notice that, the set of time series for each Temporal Abstract share the extremes of the Y axes. Besides that, the color of a time series is given by the color used on the corresponding Compact Spatial Abstract (i.e., matrix plot).

Recall that, if the value of the Spatial ANN is less than 1, the trend is toward spatial clustering, while if the value is greater than 1, the trend is toward dispersion. Very low or very high z-score values suggest some spatial pattern and, therefore, we can reject the complete spatial randomness.

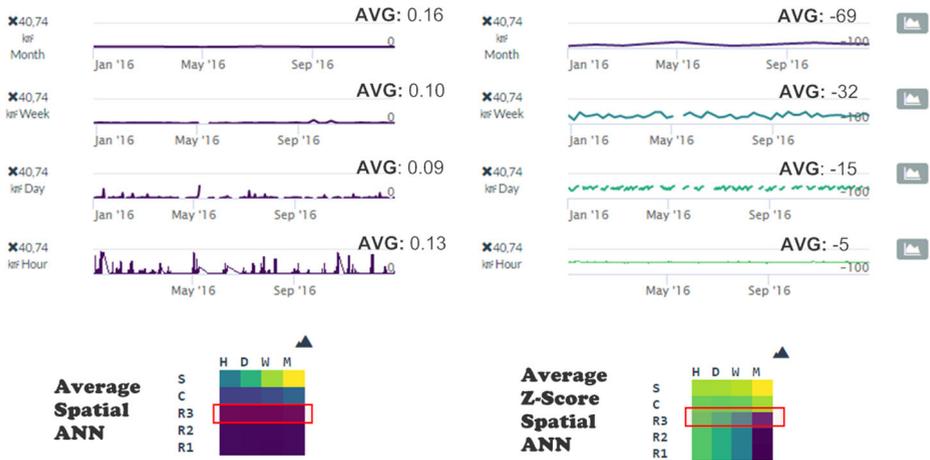


Fig. 10 The spatial average nearest neighbor and its z-score in four st-LoD

Based on Fig. 10, the Spatial Abstracts revealed a clustered phenomenon over time, since the average of the Spatial ANN values points to clusters of events throughout time. In the st-LoD (*R3, Hour*) we can observe variations between a clustered and a non clustered phenomenon. However, in the remaining st-LoD, the phenomenon reveals to be quite stable and clustered because the values of the Spatial ANN are constantly close to zero and the corresponding z-scores are quite negative (i.e., the z-score is not close to zero).

As these two Spatial Abstracts complement each other, we plot them in a scatter plot, using the interface (a click on the right-side buttons displayed in Fig. 10). These scatter plots are displayed in Fig. 11. Notice that the extremes on both axes are relative to the shown st-LoD. Each point in a scatter plot shows the values of the two Spatial Abstracts that occurred

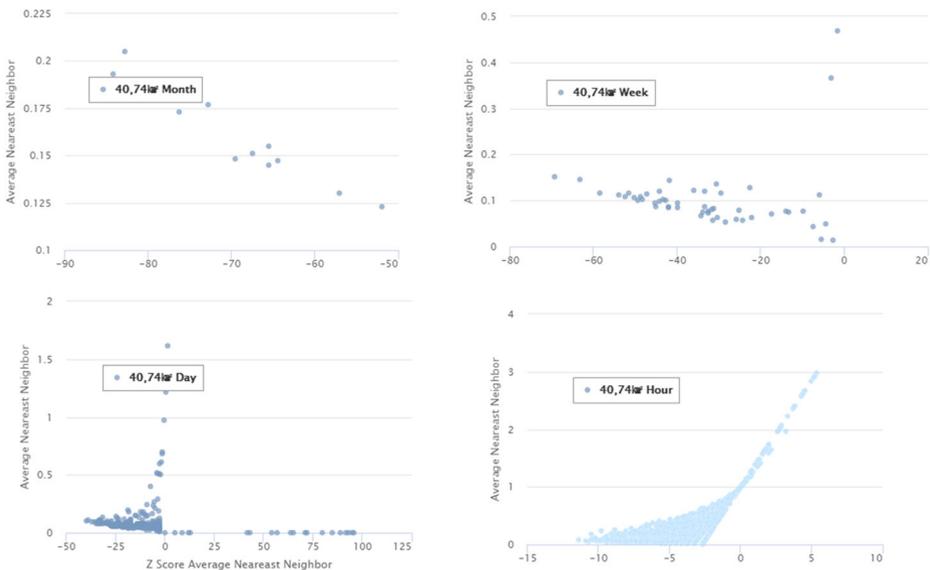


Fig. 11 The spatial ANN and its Z-score displayed in four st-LoD

at a particular temporal granule. Therefore, the number of points in a scatter plot is equal to the number of temporal granules in the temporal granularity that composes the st-LoD being displayed.

At the st-LoD (40, 74km^2 , *Hour*), there are many points holding a value close to zero in the Spatial ANN, and their z-scores are not so negative as the ones in the other st-LoDs. Looking at the st-LoD (40, 74km^2 , *Month*), it seems that the phenomenon is always clustered (i.e., the values of Spatial ANN are close to zero and their z-scores quite negative). Finally, regarding the st-LoD (40, 74km^2 , *Day*) and st-LoD (40, 74km^2 , *Week*), it seems to be the st-LoD that better fit the Poisson Cluster process. Recall that, in a Poisson Cluster process, events occur near other events but there are some points in time when no events occur. This is visible in the scatter plots of the st-LoD (40, 74km^2 , *Day*) and st-LoD (40, 74km^2 , *Week*), once the majority of the points have the values of the Spatial ANN close to zero and their z-scores are quite negative. However, there are also points where the values of the Spatial ANN are close to zero and their z-scores are positive (no clustering) and also there are points with values of the Spatial ANN that are far from zero (no clustering).

In short, the analysis made over Dataset 2 that contains a Poisson cluster process simulated with clusters of events dispersed within 110 km and one day around their parents, result in the following conclusions:

- We use the matrix plots to analyze the GMBN, *occupation rate* and *collision rate*. Here, the GMBN pointed to the st-LoD (*R3, Day*);
- We correlate the GMBN and Collision rate using the correlation of evolution through spatial granularities and through temporal granularities. Again, the st-LoD (*R3, Day*) was suggested;
- We used the phenomenon representation area to have an overview of the phenomenon at st-LoD - (*R3, Day*) in three temporal granules chosen without any particular criterion. Clusters of events were observed;
- The Temporal Abstract - Temporal Center Mass's Positioning was studied in three different LoDs. Furthermore, two Compact Temporal Abstracts were also analyzed: Coefficient of variation and the spatial autocorrelation. Here, the st-LoD suggested was also st-LoD (*R3, Day*) if one wants to understand in what periods of time clusters of events occur in certain geographic regions. It was also possible to observe that the clusters are spread out over the entire area of the USA;
- The Spatial Abstracts - spatial average nearest neighbor (Spatial ANN) and its z-score was used to understand not only when the clusters are happening but also what st-LoD better fits the Poisson Cluster process. The analysis suggested that clusters are distributed throughout the one "year" in which data was simulated. Finally, the analysis suggests that the st-LoD that better fits the Poisson Cluster process is st-LoD - (*R3, Day*) or (*R3, Week*).

Other datasets were simulated following the Poisson cluster model, namely, Datasets 3, 4 and 5. These datasets were also simulated within the USA boundaries over a year. In **Dataset 3**, each cluster of events was built around a parent within a spatial distance of 2 km and a temporal distance of one week. **Dataset 4** is similar to Dataset 2 but contains an additional set of 3.000 events following a homogeneous model. These 3.000 events are spread out over the same period of the 30.000 events that follow the Poisson Cluster model. Finally, in **Dataset 5**, each cluster of events was built around a parent within a spatial distance of 570 km and a temporal distance of one week. In the following analysis, we also add Dataset 1 that was simulated with the Homogeneous model.

The datasets described were also modeled using similar valid granularities. All the granularities are equal with respect to the previous demonstration case, except for the Raster granularities. This occurs because the minimum bounding box made by the events of the phenomenon might change from one dataset to another. Nevertheless, the most detailed spatial granularity is based on a grid of 16384 x 16384 cells and the other coarser spatial granularities were obtained by dividing the grid by a factor of 4.

Datasets 1, 3, 4, 5 will be discussed more briefly, discussing whether the *VAST* points to suitable LoDs to detail our analyses, once the “detailed” analyses would be similar to the ones made over Dataset 2. Furthermore, a comparison between the abstracts’ values obtained by a Poisson Cluster dataset or a Homogeneous dataset is made.

Figure 12 shows the global abstracts for all spatiotemporal LoDs of Datasets 1, 3, and 4. First of all, the *occupation rate* follows a similar pattern in all datasets. Dataset 3 stands out from the others regarding the Collision rate. This occurs because the clusters in Dataset 3 were simulated within a spatial distance of 2 km, so data is more spatially clustered than in the other datasets. As a result, the collision among granular syntheses starts to occur “sooner”, i.e., in finer st-LoDs when compared to the other datasets. As to Dataset 3, the GMBN highlights the following st-LoDs:

1. (*Raster*(0.1km²), *Day*);
2. (*Raster*(0.1km²), *Week*);
3. (*Raster*(1.6km²), *Day*);
4. (*Raster*(1.6km²), *Week*).

In this case, the values of spatiotemporal interaction are similar among the four st-LoDs, and therefore, any of the st-LoDs highlighted is potentially suitable to detail our analyzes.

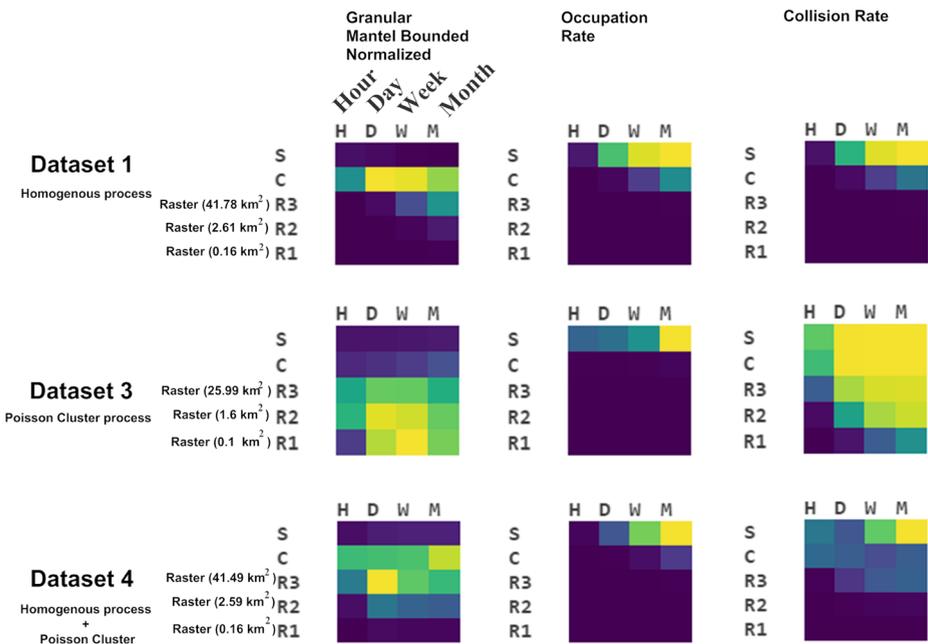


Fig. 12 Global abstracts regarding Datasets 1, 3, and 4

Nevertheless, the st-LoDs ($Raster(1.6km^2)$, $Week$) is the st-LoDs that better approaches the st-LoDs in which the data was simulated, once each cluster of events was simulated around a parent within a spatial distance of 2 km and a temporal distance of one week.

Dataset 4 is similar to Dataset 1, complemented by a homogeneous process. In this case, the GMBN suggest the st-LoDs ($Raster(41.49km^2)$, Day), which is the st-LoDs that better approaches the st-LoDs in which the pattern is simulated, once each cluster of events was simulated around a parent within a spatial distance of 110 km and a temporal distance of one day.

Nevertheless, a single Global Abstract should not be used to immediately guide our analyses for one or more st-LoDs. So far, we have been using four global abstracts to have a grasp of the data. From these four abstracts, one is neighborhood dependent (GMBN) and the remaining ones are not (*occupation rate* and *reduction rate*). In other words, only the GMBN captures the spatiotemporal dynamics of events. Therefore, restricting ourselves to just one global abstract that looks for spatiotemporal patterns or properties of the spatiotemporal interaction might wrongly suggest one or more st-LoDs as demonstrated below.

In Dataset 5, the GMBN highlights the following st-LoDs: (i) ($County$, $Hour$); (ii) ($County$, Day). However, each cluster of events was simulated around a parent within a spatial distance of 570 km and a temporal distance of one week. The problem is that the events within a cluster are spatially “dispersed” (570 km) and the GMBN is not capable of capturing such situation. In Dataset 1, the st-LoDs ($County$, Day) and ($County$, $Week$) are pointed as potential st-LoDs in which there might be spatiotemporal interaction. However, this dataset was generated following a Homogenous model. This kind of scenarios can be easily discarded when we analyze several Global Abstracts that are looking for spatiotemporal patterns, or Global Abstracts with Compact Spatial Abstracts, or Global Abstracts with Compact Temporal Abstracts, or even all together.

To illustrate the previous idea, we analyzed the correlation between the GMBN and the Average of the Spatial ANN (Compact Spatial Abstract) for the different datasets as displayed in Fig. 13.

Let’s consider Dataset 1 that is the one with the Homogeneous process. The correlation charts shows that when the GMBN reaches its maximum value, the value of the Average of the Spatial ANN is much greater than 1 (squared orange marker). Therefore, this phenomenon hardly follows a clustered pattern over time because in that case the value of the Average Spatial ANN would be closer to 0, something that did not happen in any st-LoD as the minimum value was 0.8.

In Dataset 3, we have a clear hint about the st-LoDs where the pattern was simulated because when the GMBN reaches its maximum value the Average of the Spatial ANN is close to zero (diamond green marker), as opposed to what happens in Dataset 1 (see Fig. 13).

Looking at Dataset 4, the most pronounced “elbow” tip in the chart on the left (yellow square marker) corresponds to the st-LoD ($Raster(41.49km^2)$, Day). This st-LoDs is the one that better approaches the st-LoD in which the pattern was simulated, because each cluster of events was simulated around a parent within a spatial distance of 110 km and a temporal distance of one day. Despite the fact that GMBN reaches it’s maximum in the yellow square st-LoD ($Raster(41.49km^2)$, Day), the value for the Average of Spatial ANN is 0.5 which makes the hint weaker than in the case of Dataset 3. However, this gives us a clue for the right st-LoDs.

Regarding Dataset 5, we do not have a clear hint about the st-LoDs in which the data should be analyzed. Recall that, in this dataset, each cluster of events was simulated around a parent within a spatial distance of 570 km and a temporal distance of one

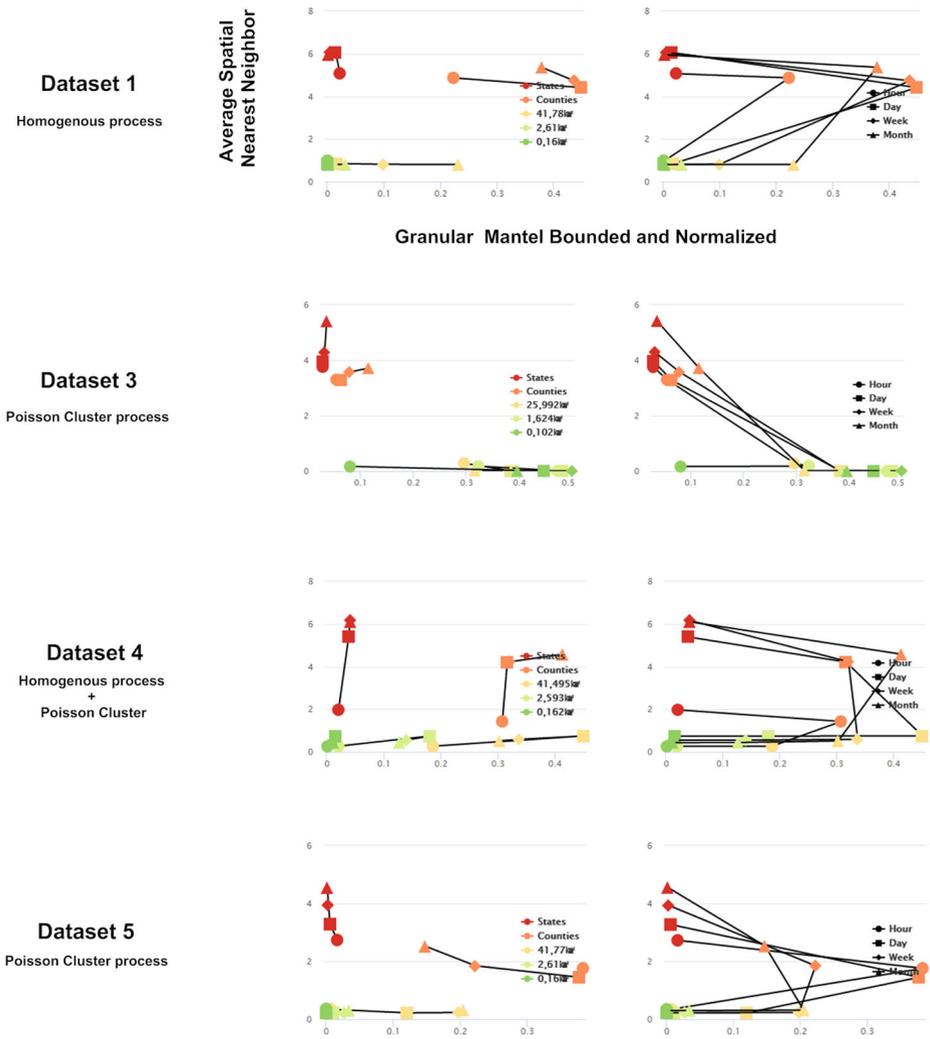


Fig. 13 Correlation between the GMBN and the average of the spatial ANN (Compact Spatial Abstract)

week. So, the events are not that clustered. Therefore, the pattern is not so pronounced when compared with the other datasets. That being said, when the GMBN reaches its maximum value the Average of the Spatial ANN is **not** close to zero (square and circle orange markers - the st-LoD (*County, Hour*) and (*County, Day*)). This result has similarities with Dataset 1 - Homogeneous process. However, in this case, two “elbow” tips are observed (i.e., st-LoD) that are not so pronounced but the Average Spatial ANN is close to zero. These correspond to the st-LoD (*Raster(41.77km²), Week*) (i.e., the diamond yellow marker) and (*Raster(41.77km²), Month*) (i.e., the triangle yellow marker). In this case, the VAST provides a hint about two st-LoD such that one of them (i.e., st-LoD - (*Raster(41.77km²), Week*)) may be appropriate to further detail the analyses. The previous analysis would not be as clear for a user that is unfamiliar with the implemented

abstracts as well as the interpretation of the provided visualizations. This relates to the learning curve concept. As a user is gaining more experience with VAST, the understanding about the concepts involved will also become clearer.

A final remark about the interpretation of the correlation charts. The “elbow” tips provide a change from a positive to a negative (or vice-versa) correlation that might be interesting to explore. Nevertheless, there might be st-LoD of interest that do not correspond to “elbow” tips. Yet, according to the values that they hold for the abstracts at study, they might be also interesting to explore as in Dataset 5.

5.1.2 Contagious process

Dataset 6 was simulated following a contagious process. The dataset was simulated within the USA boundaries over a year and is composed of 5.000 events. Based on an initial event, the next ones are generated within a spatial distance of 110 km and a temporal distance of a week. Furthermore, the dataset was modeled through the *synthetic* predicate. In this case, the most detailed spatial granularity *Raster* ($0.05km^2$) is based on a grid of 16384 x 16384 cells that covers the analyzed spatial extent of the phenomenon, and each cell has an area of $0.05km^2$. The other coarser spatial granularities were obtained by dividing the number of cells in the grid by a factor of 4. So the valid granularities for space were rasters with cell sizes approximately of $0.05km^2$, $0.8km^2$, and $12.5km^2$. The granularities *County* and *State* were also included. The time granularities used were *Hour*, *Day*, *Week*, *Month*.

To start our analysis we chose: (i) the GMBN; (ii) the Average of Spatial ANN; (iii) the Average of the z-score of the Spatial ANN; (iv) the Average of Temporal ANN; (v) the Average of the z-score of the Temporal ANN. The first three abstracts were already used so we skipped more explanations.

The Parallel Coordinates was used to simultaneously analyze the global abstracts chosen across all the st-LoDs. In this case, we are interested in understanding st-LoDs in which (i) the phenomenon seems to be more clustered over time; (ii) the phenomenon seems to be

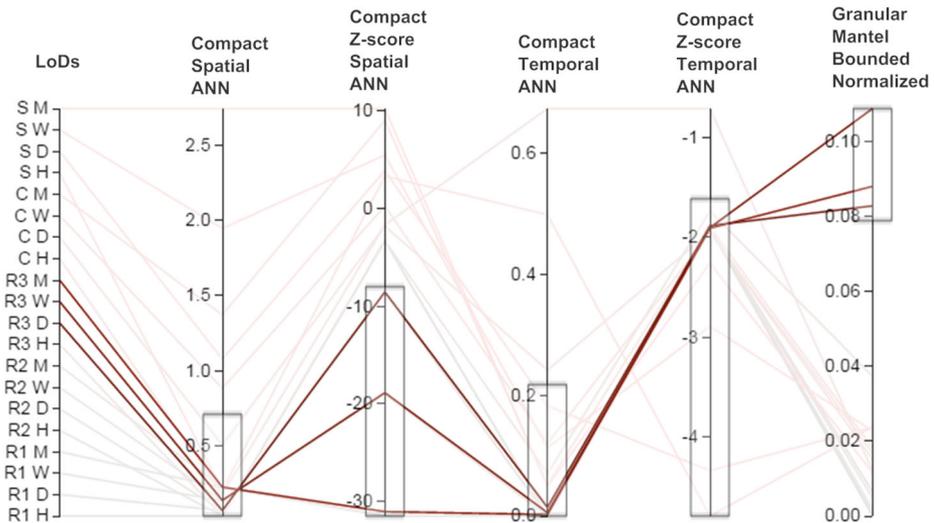


Fig. 14 Overview of the dataset 6 using global abstracts, compact spatial abstracts and compact temporal abstracts

more clustered over space; *(iii)* the st-LoDs where the spatiotemporal interaction of events seems to be better perceived. To conduct such analysis, we filtered the Parallel Coordinates in each coordinate.

This way, interactively, we just considered st-LoDs with values below 0.4 (approximately) regarding the average of the Spatial ANN. For the average of its z-score, we just considered values below -10 (approximately). Furthermore, values below 0.1 (approximately) with respect to the average of the temporal ANN were considered. For its z-score, we considered values below -1. Finally, the top three values of the GMBN were considered, which means values above 0.08. The results are displayed in Fig. 14.

Three st-LoDs were highlighted:

1. (12.5km², Day);
2. (12.5km², Week);
3. (12.5km², Month).

Like it was done in Dataset 2, the Temporal Center Mass’s Positioning was used to relate geographic regions with the center’s of mass of time at which events happened. This Temporal Abstract is displayed in Fig. 15 for the three st-LoDs identified.

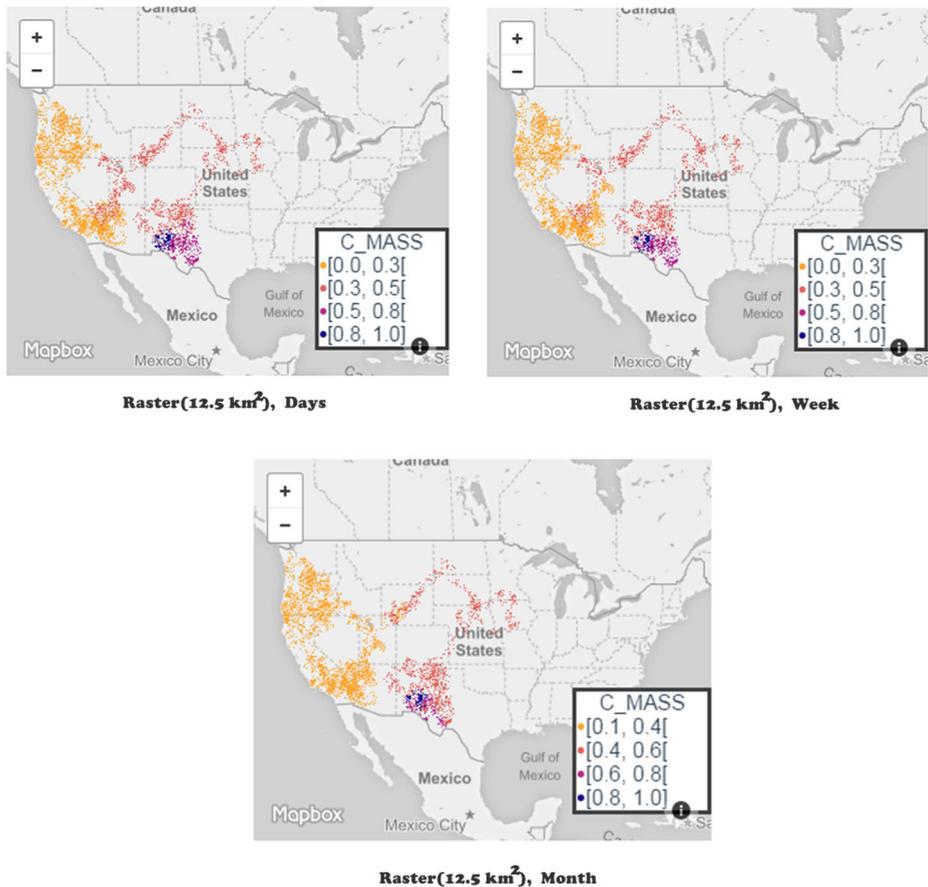


Fig. 15 One temporal abstract at three different st-LoDs

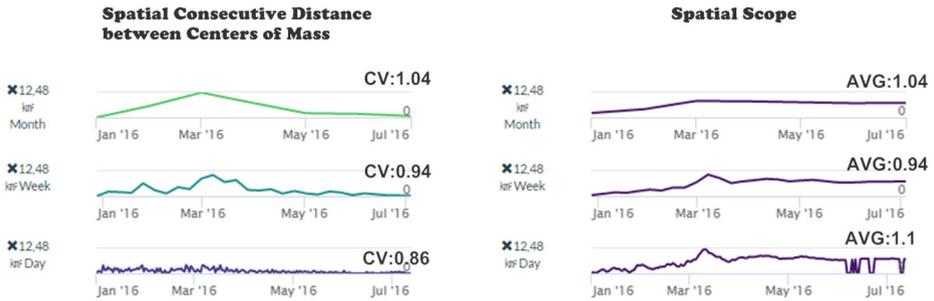


Fig. 16 Two spatial abstracts about dataset 6

Regardless of the st-LoD, a grasp about the spatial path made by the simulated contagious process is visible, thus confirming a contagious process. Nevertheless, in st-LoD (12.5km^2 , *Day*) is where the path is slightly better perceived.

Another experiment was made with two Spatial Abstracts: (i) the Spatial Scope; (ii) the Spatial Consecutive Distance between Centers of Mass. The former abstract indicates how much a phenomenon changes the size of its spatial extent over time while the latter measure whether such spatial extent moves in space over time. For the st-LoD identified initially, the Spatial Abstracts can be seen in Fig. 16. Moreover, in the former abstract the average value is displayed while in the latter the coefficient of variation is shown.

Let us start by the Spatial Scope. In general, for the identified st-LoD, the phenomenon's spatial scope is quite stable throughout time with some variations here and there. However, the most stable st-LoD is (*Raster*(12.5km^2), *Month*).

Regarding the Spatial Consecutive Distance between Centers of Mass, the st-LoD (*Raster*(12.5km^2), *Day*) is where the distances between centers of mass seem to vary less according to the coefficient of variation. Thus, if we are interested in understanding how the contagious process evolved, in this simulated scenario, we should look at the st-LoD (*Raster*(12.5km^2), *Day*) because this is the st-LoD that seems to capture the smoothest transitions over time.

To conclude, in the Contagious process an initial event is generated and, then, the next events are simulated within a specified spatial and temporal distance. The dataset under analysis was generated with distances of 110 km and one week. The events generated within neighborhood are uniformly distributed and they are not necessary at a distance of a week. In fact, many of them might be at temporal distance less than a week. This might explain why the Contagious process seems to be better perceived in the st-LoD (*Raster*(12.5km^2), *Day*).

5.1.3 Log-Gaussian Cox process

Dataset 7 (whose results are illustrated in Fig. 17) was simulated following the Log-Gaussian cox process. The dataset was simulated within the USA boundaries over a year and is composed by 15.000 events. Therefore, this dataset will show geographic regions of higher incidence of events over others.

In this case, the most detailed spatial granularity R1 is based on a grid of 16384 x 16384 cells that covers the analyzed spatial extent of the phenomenon, and each cell has an area of 0.16km^2 . The remaining valid raster granularities for space were 2.57km^2 and 41.27km^2 .

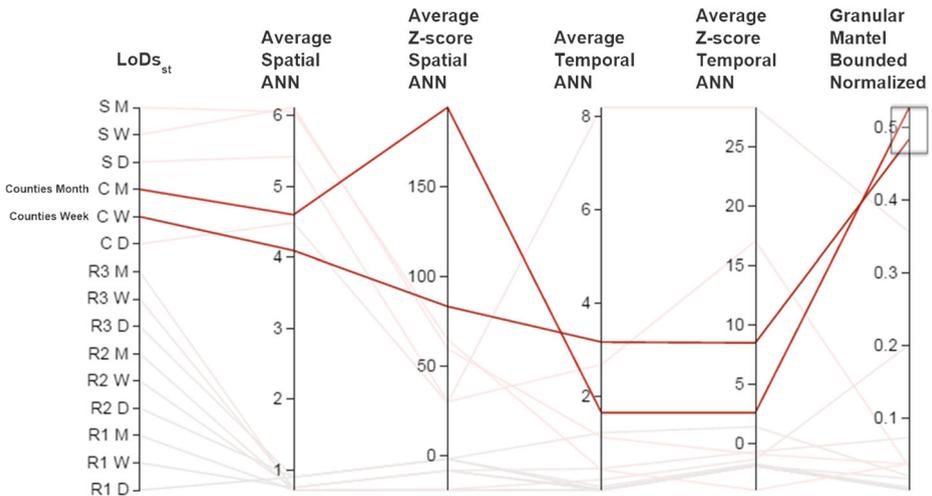


Fig. 17 Overview of the dataset 7 (Log-Gaussian cox process) using global abstracts, compact spatial abstracts and compact temporal abstracts

The granularities *County* and *State* were also included. The time granularities used were *Day*, *Week*, *Month*.

As in Dataset 6 (contagious process), we start by getting an overview of the set of the following abstracts using the Parallel Coordinates: (i) the GMBN; (ii) the average of Spatial ANN; (iii) the average of the z-score of the Spatial ANN; (iv) the average of temporal ANN; (v) the average of the z-score of the temporal ANN. Looking at the Parallel Coordinates:

- There are no st-LoDs holding values close to zero with respect to Average Spatial ANN, containing also quite negative z-scores. This kind of values suggest that we are not dealing with the Poisson cluster process as events occur close to each other in space.
- There are some st-LoDs holding values close to zero with respect to Average Temporal ANN but their z-scores are also close to zero. Also, for such cases, the spatiotemporal interaction is weak when compared with other st-LoDs. These values suggest that we are not dealing with the contagious process as events occur close to each other in space and in time.
- Two st-LoDs have the spatiotemporal interaction among events measured by the GMBN above 0.4, which is similar to the values obtained in Poisson Cluster simulated datasets. However, at this point, no particular meaning can be assigned to such values.

In Log-Gaussian Cox processes, we have geographic regions of higher incidence that might change slowly over time. This way, there are geographic regions that distinguish themselves from others in terms of the number of events that happened in there, as well as the geographic regions of higher incidence that might “infect” their neighbours. Since Log-Gaussian Cox processes simulate geographic regions of higher incidence, temporal abstracts might be useful. Hence, we chose the **Temporal Frequency Rate** that measures, for each spatial granule, the percentage of atoms that occurred on it, given all the atoms of the phenomenon at a given LoD. To capture the st-LoD where the Log-Gaussian Cox process is better perceived, we correlate the Compact Temporal Abstract - Coefficient of variation and the Spatial autocorrelation of the temporal frequency rate as can be seen in Fig. 18. These two Compact Temporal Abstracts are chosen because we want to capture

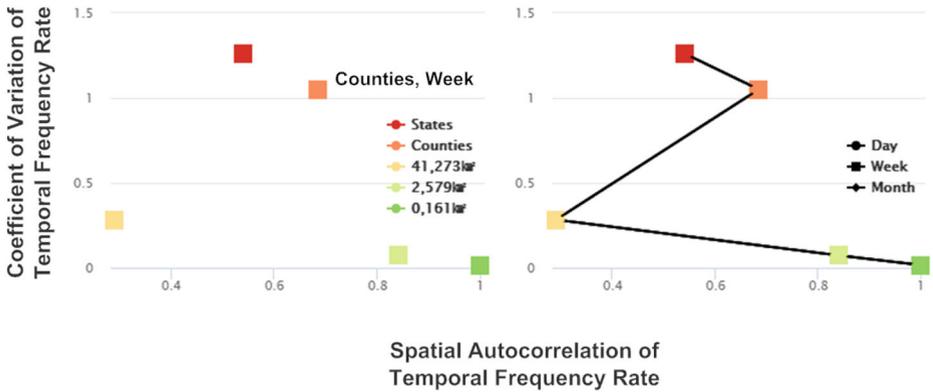


Fig. 18 Dataset 7 (Log-Gaussian Cox process) - correlation between the coefficient of variation of temporal frequency rate and the spatial autocorrelation of temporal frequency rate

the st-LoDs in which there is a considerable variation on the Temporal frequency rate, and simultaneously, to understand whether the spatial granules are spatially correlated on the Temporal frequency rate.

First of all, the temporal granularity does not have an impact on the **Temporal Frequency Rate**. Regardless the temporal granularity, the percentage of atoms that occurred on particular spatial granules remains the same, as can be observed on the left chart of Fig. 18.

That being said, let us focus on the right chart in Fig. 18. In finer spatial granularities, a spatial autocorrelation among spatial granules is expected to exist, since their values should

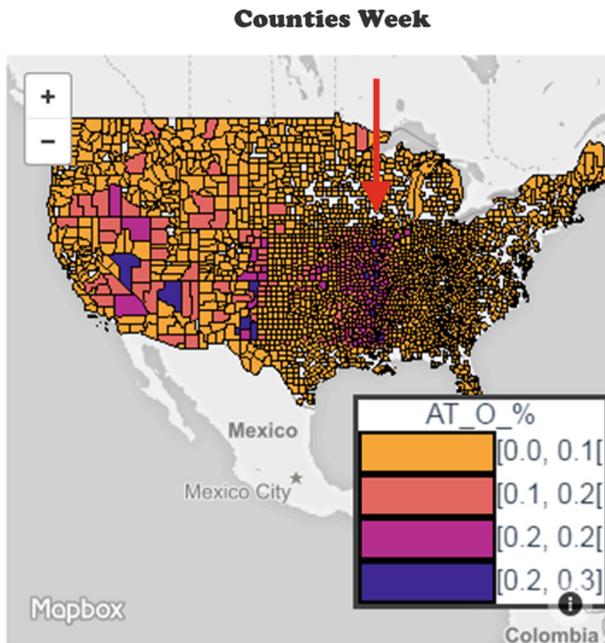


Fig. 19 The temporal frequency rate at the st-LoD (*County, Week*)

diverge little or nothing as shown by their coefficient of variation. But when we look at the st-LoD (*County, Week*), the coefficient of variation is a value near to one, which indicates variability among values, and simultaneously, the level of spatial autocorrelation grows. But if we move to st-LoD (*State, Week*), the spatial autocorrelation decreases. To check the previous analysis, the Temporal Frequency Rate is shown in Fig. 19 at the st-LoD (*County, Week*). There are some counties (that are spatially small) on the east side of USA (highlighted with a red arrow) that have a higher incidence of events. The corresponding geographic area was zoomed-in and displayed at two st-LoDs: (i) (*County, Week*); (ii) (*Raster (41.27km²), Week*) as shown in Fig. 20.

Looking at the st-LoD (*Raster (41.27km²), Week*) geographic regions with higher incidence of events are no longer perceived. Although there are geographic regions with higher incidence (purple and dark blue spatial granules), the values of the **Temporal Frequency Rate** are not as different as they are in the st-LoD (*County, Week*). This confirms that the st-LoD (*County, Week*) is probably one of the suitable st-LoDs to better understand the geographic regions that are most affected by the phenomenon.

5.2 Real datasets

Several phenomena were analyzed using VAST. As opposed to synthetic datasets, we are not aware of possible patterns that those phenomena might contain.

5.2.1 Wildfires in Portugal

In this section, we report the analysis made about wildfires that occurred in Portugal between 2001 and 2012.

The granularities-based model was used to model them at different LoDs. This phenomenon is described by a collection of 280.968 spatiotemporal events. These events were modeled through the wildfires predicate containing two arguments: *wildfires(space, time)*. The following pages shows the analysis made about wildfires that occurred in Portugal between 2001 and 2012. The most detailed spatial granularity *Raster (0.005km²)* is based on a grid of 16384x 16384 cells that covers the analyzed spatial extent of the phenomenon, and each

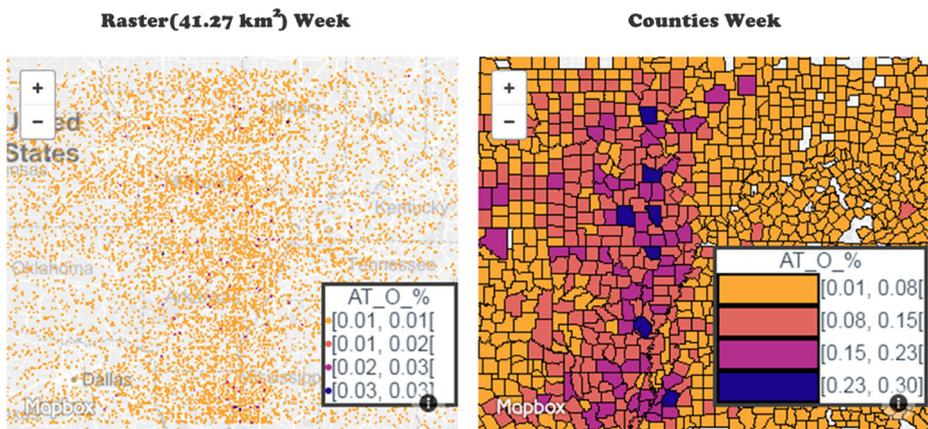


Fig. 20 The **Temporal Frequency Rate** at the st-LoD (*Raster (41.27km²), Week*) and (*County, Week*)

cell has an area of 0.005 km^2 . The remaining raster granularities for space were granularities with cell sizes of 0.08 km^2 and 0.319 km^2 . The granularities *Parish*, *County* and *District* were also considered. The time granularities used were *Hour*, *Day*, *Week*, *Month*, *Year*.

The considered granular terms required to model these events were: *Instant* and *Cell* for the time and space arguments, correspondingly. The raw data was encoded at the finer st-LoD, which includes the time granularity of *Hour* and the space granularity *Raster* (0.005 km^2).

To have a grasp of wildfires in Portugal, we chose the following abstracts: (i) the GMBN; (ii) the average of Spatial ANN; (iii) the average of the z-score of the Spatial ANN; (iv) the average of Temporal ANN; (v) the average of the z-score of the Temporal ANN. Parallel Coordinates were used to simultaneously analyze the global abstracts chosen across all the st-LoDs (see Fig. 21). Let us now take a close look at them:

- There are st-LoDs holding values close to zero with respect to Average Spatial ANN, that simultaneously have quite negative values considering the corresponding z-score. Therefore, these values has some resemblance with the ones obtained with the Poisson cluster simulated datasets, or, with the contagious ones. As a result, at this point, we might say that wildfires in Portugal hardly follow a homogeneous model;
- Several st-LoDs are holding values close to zero with respect to Average Temporal ANN but their z-scores are close to zero, which means that the complete randomness cannot be rejected. In other words, wildfires occurring on the same spatial granule are likely not close to each other in time, on average. Furthermore, this information is telling us that, probably, we are not dealing with a phenomenon that follows a Contagious process;
- Several st-LoDs have the spatiotemporal interaction among events measured by the GMBN above 0.4, which is similar to the values obtained in the Poisson Cluster simulated datasets. This reinforces the similarities of the wildfires in Portugal with the Poisson Cluster model.

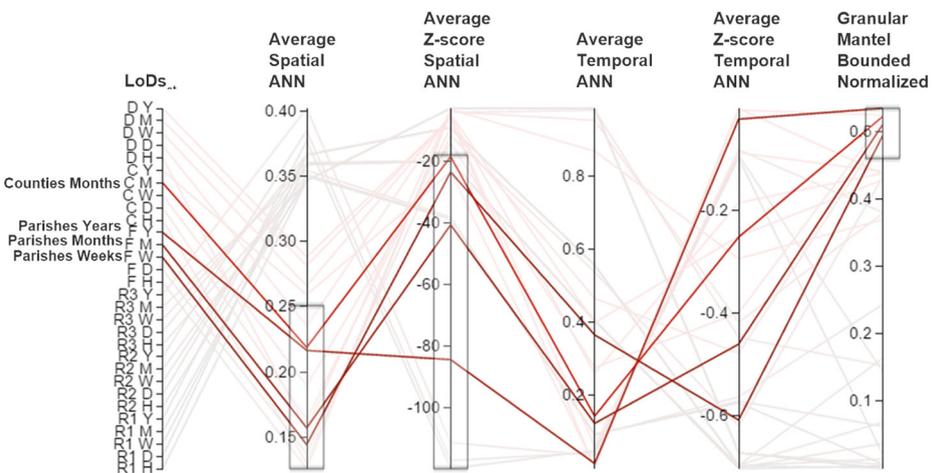


Fig. 21 Overview of wildfires in Portugal using global abstracts, compact spatial abstracts and compact temporal abstracts

The Parallel Coordinates visualization was filtered to identify the suitable st-LoDs to confirm the previous hypothesis. We just considered st-LoDs with values below 0.25 (approximately) regarding the Average of the Spatial ANN. For the average of the corresponding z-score, we just considered values below -20 (approximately). Finally, the top four values of the GMBN were considered, which means values above 0.45 (approximately). The other coordinates (temporal average nearest neighbor and its z-score) were not filtered because there are no domain values that clearly point to clustered or dispersed events in time. From the filtering, four st-LoDs were highlighted: (*Parish, Week*), (*Parish, Month*), (*Parish, Year*), (*County, Month*).

To better understand how wildfires occur in space over time, the Spatial ANN and its z-score were plotted in a scatter plot for these st-LoDs (Fig. 22). First of all, notice that the charts present similarities in the values and corresponding “shapes” with the charts obtained when we studied Poisson Cluster simulated datasets. Furthermore, in all the displayed st-LoDs, the phenomenon reveals to have several clustered distributions of events over time.

Nevertheless, st-LoDs (*Parish, Week*) is the one that better fits the Poisson Cluster process/model. That means, in general, that events occur near one another but there are a few times when events did not occur or occur in a dispersed way. Furthermore, in this st-LoD there is a good trade-off between the Spatial ANN and its z-score. In other words, there are many temporal granules in which the Spatial ANN’s values are around 0.15 (trend toward clustering) and where their z-scores are quite negative (confirmation of clustering).

VAST allows users to zoom-in on a particular area of a scatter plot. When that action is performed, the selected points (i.e., temporal granules) are highlighted on the corresponding time-series using vertical red lines. Thus, to understand when wildfires are occurring spatially clustered, we zoom-in the scatter plot at st-LoDs (*Parish, Week*) over the area where the Spatial ANN is less than 0.2 and its z-score is less than -35. We are choosing

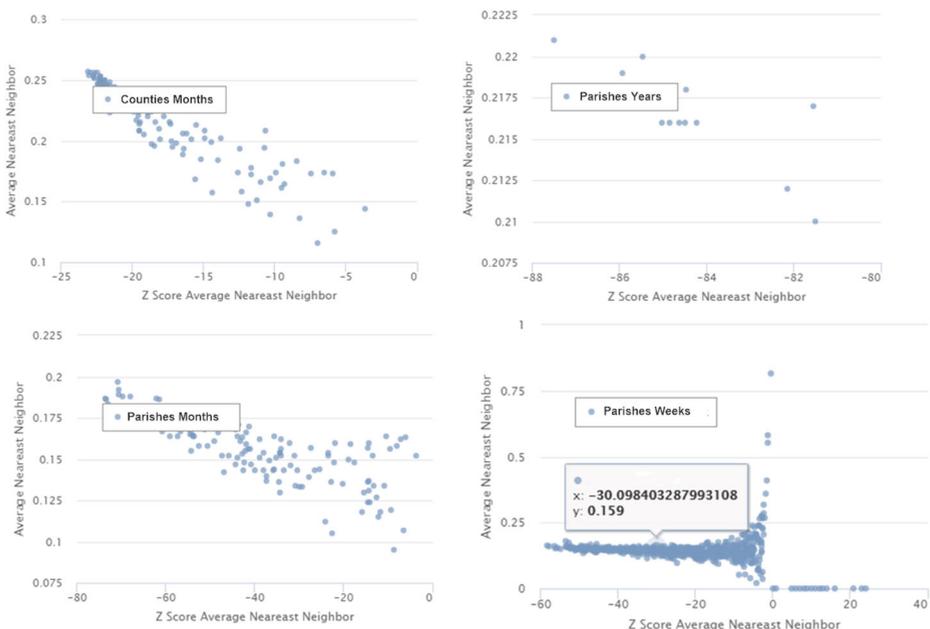


Fig. 22 The Spatial ANN and its Z-score displayed in four st-LoDs

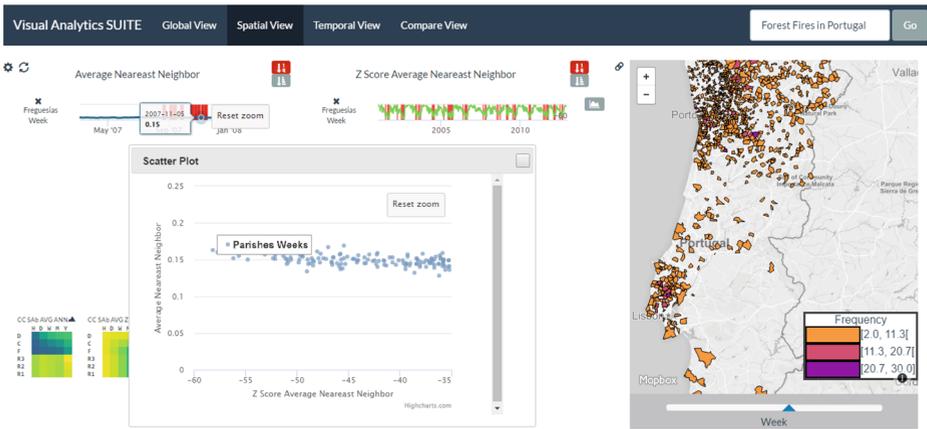


Fig. 23 Filter the temporal granules in which the clusters of events are most pronounced at st-LoDs (*Parish, Week*)

the temporal granules in which the events are more spatially clustered. The result of this selection can be seen in Fig. 23.

The time series on the right-hand side is showing the entire temporal period at study. From it, we can notice that the wildfires occurred recurrently spatially clustered which, in general, matches the summer periods but not necessarily. For instance, during the week that has started on November, 5th 2011, several wildfires occurred in Portugal. These are displayed on the map of Fig. 23, and it is possible to confirm that they are spatially clustered.

5.2.2 Violence against civilians

This section shows the analysis made over the dataset concerning violence against civilians in Africa that occurred between 1997 and 2015. The granularities-based model was used to model the events at different LoDs. This phenomenon is described by a collection of 33.393 spatiotemporal events. These events were modeled through a terrorism predicate, with two arguments *terrorism(space, time)*.

The most detailed spatial granularity *Raster* (343.45km^2) is based on a grid of 16384×16384 cells that covers the analyzed spatial extent of the phenomenon, and each cell has an area of 343.45 km^2 . The other coarser spatial granularities were obtained by dividing the number of cells in the grid by a factor of 2. So the valid granularities for space were rasters with cell sizes of 1376.34 km^2 , 5525.79 km^2 , and 22268.15 km^2 . The used time granularities were *Hour, Day, Week, Month, Year*.

Like previously, we started figuring out what kind of model might be underlying this phenomenon, using the usual abstracts: (i) the GMBN; (ii) the Average of Spatial ANN; (iii) the Average of the z-score of the Spatial ANN; (iv) the Average of Temporal ANN; (v) the Average of the z-score of the Temporal ANN. The Parallel Coordinates are displayed in Fig. 24, and the main findings are:

- There are “four levels” of spatial clustering over time, as depicted by the Average Spatial ANN and the corresponding z-scores. These levels are being strongly influenced by the temporal granularity. With the temporal granularity *Year*, the Average Spatial ANN

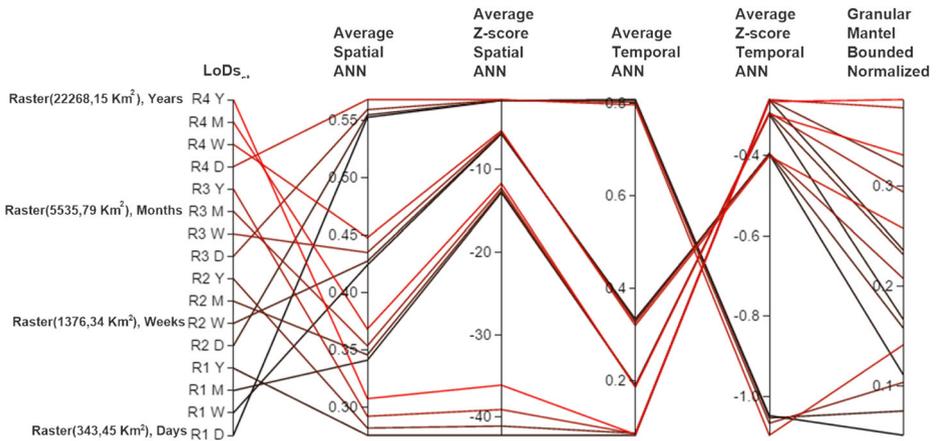


Fig. 24 Overview of the attacks against civilians in Africa using global abstracts, compact spatial abstracts and compact temporal abstracts

and the corresponding z-scores reach their minimums, while with the temporal granularity *Day* the spatial clustering over time is not so pronounced. Thus, the phenomenon seems to have some similarities to the Poisson Cluster model;

- Several st-LoDs are holding values close to zero with respect to the Average Temporal ANN but their z-scores are also close to zero, which means that the complete randomness cannot be rejected. In other words, the attacks against civilians occurring on the same spatial granule are likely not close to each other in time, on average. Furthermore, this information is telling us that we are likely not dealing with a phenomenon that follows a Contagious process. This is quite similar to the phenomenon about Wildfires in Portugal;
- There are some st-LoDs that have the spatiotemporal interaction among events measured by the GMBN above 0.3, which is similar to the datasets simulated with the Poisson cluster or with the dataset about wildfires in Portugal.

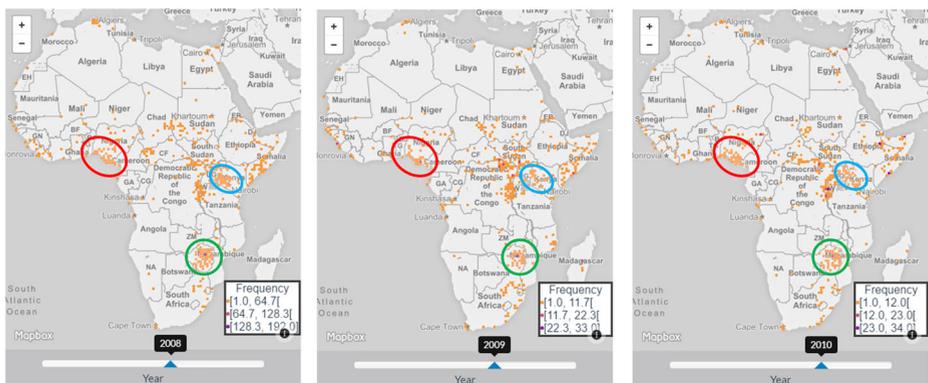


Fig. 25 Violence against civilians at the st-LoD $Raster(22268.15\ km^2)$, Year displayed in three temporal granules - 2008, 2009 and 2010

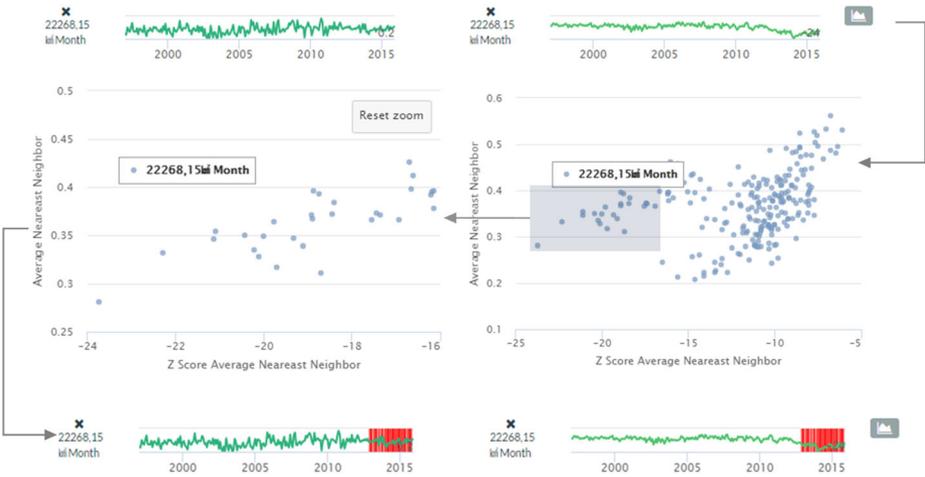


Fig. 26 Highlighting the temporal granules where the violence against civilians is more spatially clustered using the st-LoD ($Raster(22268.15\text{ km}^2, \text{Month})$)

Since the Average Spatial ANN and the corresponding z-scores reach their minimums, we look to the data at the st-LoD ($Raster(22268.15\text{ km}^2, \text{Year})$) in three temporal granules: 2008, 2009, 2010. The temporal granules were chosen for no particular reason, to try seeing if there were clusters of events based on the tip provided by the Parallel Coordinates (see Fig. 24).

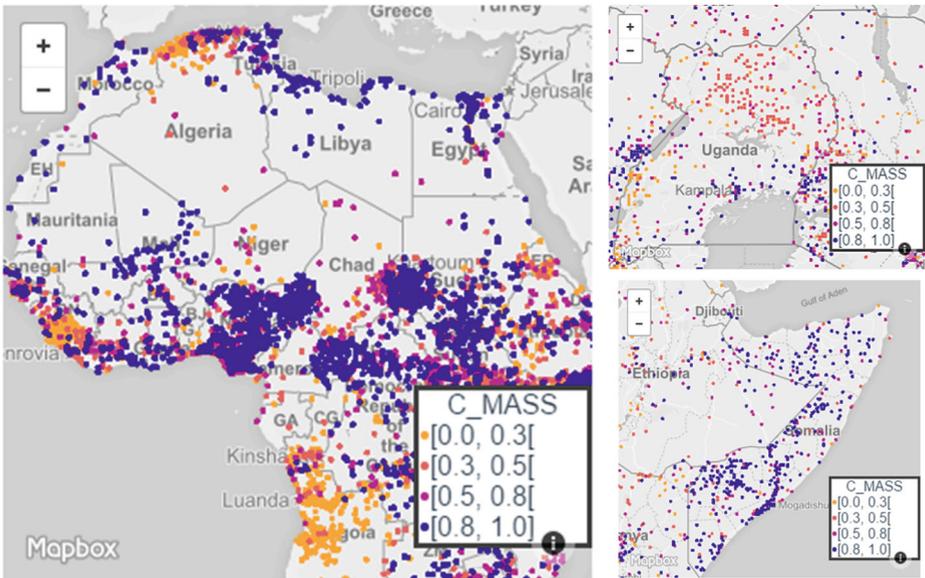


Fig. 27 Violence against civilians at the st-LoD $Raster(343.45\text{ km}^2, \text{Week})$ displayed in three different spatial extents

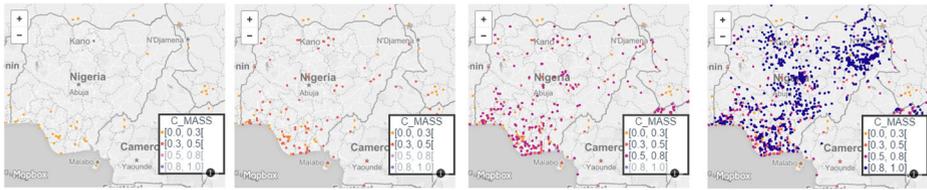


Fig. 28 Evolution of violence against civilians throughout time at the st-LoDs *Raster*(343.45 km², *Week*) in Nigeria

As we can see in Fig. 25, the spatiotemporal events are in fact spatially clustered. In this case, there are clusters of events that remain stable in the three years chosen, like for example the cluster in Mozambique (green circle), South Nigeria (red circle), and on the border between Uganda and Kenya (blue circle). In our initial analysis about violence against civilians, the st-LoDs containing the temporal granularity *Month* also suggest some characteristics of the Poisson Cluster process and, consequently, clusters of events over time. So, we have chosen the st-LoD (*Raster*(22268.15 km²), *Month*) for displaying the Spatial ANN and the corresponding z-score. Afterward, we plot them in a scatter plot and filter out the temporal granules where the values of Spatial ANN are low and the values of the z-score are more negative, that is, the temporal granules where the clusters are likely most pronounced. This action highlights the time series on the respective granules as displayed in Fig. 26.

Surprisingly, only “recent” temporal granules were highlighted, which means that the attacks against civilians in Africa are getting more spatially clustered than in the past. A more detailed analysis about this change can be done and the st-LoDs (*Raster*(22268.15 km²), *Month*) seems suitable for such an analytical task.

Another experiment was made to understand whether the attacks against civilians occur on the same regions over time or if there were changes. To conduct this analysis, we have chosen the Temporal Center Mass’s Positioning. Since the spatial autocorrelation of the Temporal Center Mass’s Positioning is most pronounced in st-LoDs containing the spatial granularity *Raster*(343.45 km²), our analysis was detailed in the st-LoD *Raster*(343.45 km², *Week*). Some results are displayed in Fig. 27.

Clusters of events are changing over time. For instance, in Angola most of the attacks occurred in the past and they are not that frequent anymore. The same is observed in Serra Leoa. But for instance, at northern Algeria, the attacks have slightly changed the location over time from north-west to north-east. Looking at the north of Uganda (top right-hand map), there is no particular pattern. In each spatial granule there may be old and recently attacks, or instead, the attacks happened somewhere in the middle of the period under study (1997–2015). At Somalia (bottom right-hand map), most attacks are recent and spread out by the entire country.

In the case of Nigeria, we have used a particular feature of *VAST* that allow users to hide and show the events holding a particular class of values. In Fig. 28, from left to right, the classes were incrementally added to the map. As you can see, in the past, most attacks occurred at the south of Nigeria and, afterward, they started to spread across the entire country.

6 Conclusions and future work

The LoD plays a crucial role when analysing spatiotemporal data. From one LoD to another, some patterns can be perceived more easily, or different patterns may be detected. Modeling

phenomena at different LoDs is needed, as there is no single LoD at which data can be analyzed.

Current practices work mainly on a single LoD, driven by the analysts' perception and ignoring the fact that identifying the suitable LoDs is key for pointing relevant patterns. To enhance the analyses over spatiotemporal events, we propose to move from a single user-driven LoD to a multiple LoDs analysis approach, providing the user with an understandable high-level overview of the underlying structure of the phenomenon for each LoD. This approach can give several hints about different facets of spatiotemporal events, providing a first insight on the presence or absence of patterns at particular LoDs.

VAST was developed to support analyses in this new mindset. The tool allows to visually inspect hints about the absence or presence of different spatiotemporal patterns at multiple LoDs, simultaneously, following a coordinated and integrated strategy among the provided visualizations.

The VAST tool is based on the Visualization of Abstracts, proposed within the scope of the SUITE framework, using an integrated set of visualizations including: parallel coordinates and matrix plots for the ST-Abstracts; linked temporal series for the Spatial Abstracts; linked maps for the Temporal Abstracts; a linked map with temporal sliders for the real data phenomena; and scatters for the correlation analysis for pairs of abstracts. In this paper we have demonstrated how the joint use of these visualizations allows the detection of many spatiotemporal patterns in the data, and at which LoDs they are better perceived. We also proposed to use a new abstract, namely the Granular Mantel Bounded and Normalized (GMBN) abstract, that measures the spatiotemporal interaction among granular syntheses. This abstract was very important in the process of patterns discovery.

To the best of our knowledge, there is no other prototype or application that supports the analysis of spatiotemporal events at multiple LoDs, simultaneously, following the VA Mantra.

Experiments were conducted with two types of datasets describing spatiotemporal events, namely: (i) synthetic datasets and (ii) real datasets. Synthetic datasets with different spatiotemporal patterns (Poisson Cluster Process, Contagious Process, or Log-Gaussian Cox Process) at different LoDs were produced. For most cases, VAST could provide a correct overview of the phenomenon, allowing us to identify the LoDs where patterns exist and, therefore, the LoDs that should be used to detail the analysis. Two real datasets were also analyzed and discussed: (i) wildfires that occurred in Portugal between 2001 and 2012; and, (ii) violence against civilians in Africa that occurred between 1997 and 2015. In general, VAST was able not only to give an overview of the presence or absence of different spatiotemporal patterns, but also to suggest the proper spatiotemporal LoDs that allow us to better perceive the corresponding patterns.

Future work can be directed towards the development of heuristics to automatically suggest LoDs to analyze the data. These heuristics are needed and should be a priority if the number of abstracts grows considerably, as manual selection might be overwhelming to the user. This issue relates to the learning curve. Each abstract looks for a feature or pattern which frequently is expressed in terms of a range of values. According to the value, it means one thing or the other. Thus, a user needs to get familiar with the abstracts and their interpretation. Requiring a user to memorize all the abstracts and their interpretation might be overwhelming, specially if we consider the joint interpretation of abstracts. Once again, heuristics to automatically suggest LoDs are needed.

Acknowledgments This work has been supported by FCT - Fundação para a Ciência e Tecnologia MCTES, UID/CEC/04516/2013 (NOVA LINCS), UID/CEC/00319/2019 (ALGORITMI), and UID/CEC/50021/2019 (INESC-ID).

Appendix : Abstracts implemented

Measure	Description	Global	Spatial	Temporal
BRAY-CURTIS SIMILARITY FOR SYNTHESIS	Calculates the similarity based on the number of granular synthesis, between consecutive temporal grains		•	
CORRELATION INDEX FOR ATOMS	Correlation between the number of atoms of consecutive temporal grains		•	
CORRELATION INDEX FOR SYNTHESIS	Correlation between the number of granular synthesis of consecutive temporal grains		•	
DICE SIMILARITY (BINARY)	Dice index (event / no event) between consecutive temporal grains		•	
JACCARD SIMILARITY (BINARY)	Jaccard index (event / no event) between consecutive temporal grains		•	
GOWER SIMILARITY (BINARY)	Similarity (event / no event) between consecutive temporal grains		•	
MORAN'S I	Calculates the spatial autocorrelation among nearby locations, given a domain specific variable		•	
NEAREST NEIGHBOR (NN)	Measures the level of clustering		•	•
Z-SCORE NEAREST NEIGHBOR (Z-NN)	Measures the z-score of the level of NN		•	•
SPATIAL SCOPE	Measures the spatial extent		•	
SPATIAL CONSECUTIVE DISTANCE BETWEEN CENTERS OF MASS	Measures the distance between consecutive centers of mass		•	
CENTER'S MASS POSITIONING	Measures the position of the centers of mass		•	•
REDUCTION RATE (%)	Measures the reduction of atoms used	•		
AVERAGE ATOMS IN SPATIOTEMPORAL GRANULES (%)	Measures the average of atoms indexed by spatio-temporal granules	•		

COLLISION RATE (%)	Percentage of spatiotemporal granules with events, where atom collisions exists	•	•	•
OCCUPATION RATE (%)	Percentage of granules with events	•	•	•
GRANULAR MANTEL BOUNDED AND NORMALIZED	Measures the spatio-temporal interaction	•		
FREQUENCY RATE (%)	Percentage of events happened in granules		•	•
BRAY-CURTIS SIMILARITY FOR ATOMS	Calculates the similarity based on the counts of atoms, between consecutive temporal grains		•	

References

- Aigner W, Miksch S, Schumann H, Tominski C (2011) Visualization of time-oriented data. Springer Science & Business Media
- Andrienko G, Andrienko N, Bak P, Keim D, Kisilevich S, Wrobel S (2011) A conceptual framework and taxonomy of techniques for analyzing movement. *J Vis Lang Comput* 22(3):213–232
- Andrienko G, Andrienko N, Bosch H, Ertl T, Fuchs G, Jankowski P, Thom D (2013) Thematic patterns in georeferenced tweets through space-time visual analytics. *Comput Sci Eng* 15(3):72–82
- Andrienko N, Andrienko G (2004) Interactive visual tools to explore spatio-temporal variation. In: Proceedings of the working conference on advanced visual interfaces. ACM, pp 417–420
- Andrienko N, Andrienko G (2006) Exploratory analysis of spatial and temporal data: a systematic approach. Springer, Berlin
- Bédard Y, Rivest S, Proulx MJ (2007) Spatial. online analytical. processing (solap): concepts, architectures, and solutions. *Data warehouses and OLAP: concepts, architectures, and solutions*, Idea Group Inc, pp 298–319
- Bertin J, Berg WJ, Wainer H (1983) *Semiology of graphics: diagrams, networks, maps*, vol 1. University of Wisconsin press Madison
- Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) *Time series analysis: forecasting and control*, Wiley, New York
- Cardoso D, Alves R, Pires JM, Birra F, Silva R (2017) Gisplay-extensible web api for thematic maps with webl. In: International conference on computational science and its applications. Springer, pp 674–689
- Chae J, Thom D, Bosch H, Jang Y, Maciejewski R, Ebert DS, Ertl T (2012) Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In: 2012 IEEE conference on visual analytics science and technology (VAST). IEEE, pp 143–152
- Chen H, Chung W, Xu JJ, Wang G, Qin Y, Chau M (2004) Crime data mining: a general framework and some examples. *Computer* 37(4):50–56
- Cho I, Dou W, Wang DX, Sauda E, Ribarsky W (2016) Vairoma: a visual analytics system for making sense of places, times, and events in roman history. *IEEE Trans Vis Comput Graph* 22(1):210–219
- Dykes J, MacEachren A, Kraak M (2005) Exploring geovisualization. No vol 1 in International Cartographic Association. Elsevier, Amsterdam
- Ebdon D (1985) *Statistics in geography*. Blackwell, Oxford
- Ferreira N, Poco J, Vo HT, Freire J, Silva CT (2013) Visual exploration of big spatio-temporal urban data: a study of new york city taxi trips. *IEEE Trans Vis Comput Graph* 19(12):2149–2158
- Forlines C, Wittenburg K (2010) Wakame: sense making of multi-dimensional spatial-temporal data. In: Proceedings of the international conference on advanced visual interfaces. ACM, pp 33–40
- Fuchs G, Schumann H (2004) Visualizing abstract data on maps. In: 2004 Proceedings. Eighth international conference on information visualisation, 2004. IV. IEEE, pp 139–144

18. Gabriel E (2014) Estimating second-order characteristics of inhomogeneous spatio-temporal point processes. *Methodol Comput Appl Probab* 16(2):411–431
19. Gabriel E, Rowlingson B, Diggle P (2013) stpp: an r package for plotting, simulating and analyzing spatio-temporal point patterns. *J Stat Softw* 53(2):1–29
20. Gao Y, Wang S, Padmanabhan A, Yin J, Cao G (2018) Mapping spatiotemporal patterns of events using social media: a case study of influenza trends. *Int J Geogr Inf Sci* 32(3):425–449
21. Gatalsky P, Andrienko N, Andrienko G (2004) Interactive analysis of event data using space-time cube. In: 2004 Proceedings. Eighth international conference on information visualisation, 2004. IV. IEEE, pp 145–152
22. Getis A (1992) The analysis of spatial association by use of distance statistics. *Geogr Anal* 24(3):189–206
23. Goodwin S, Dykes J, Slingsby A, Turkay C (2016) Visualizing multiple variables across scale and geography. *IEEE Trans Vis Comput Graph* 22(1):599–608
24. Guo D, Chen J, MacEachren AM, Liao K (2006) A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE Trans Vis Comput Graph* 12(6):1461–1474
25. Hadlak S, Tominski C, Schulz HJ, Schumann H (2010) Visualization of attributed hierarchical structures in a spatiotemporal context. *Int J Geogr Inf Sci* 24(10):1497–1513
26. Hering AS, Bell CL, Genton MG (2009) Modeling spatio-temporal wildfire ignition point patterns. *Environ Ecol Stat* 16(2):225–250
27. Jacquez GM (1996) A k nearest neighbour test for space–time interaction. *Stat Med* 15(18):1935–1949
28. Kapler T, Wright W (2005) Geotime information visualization. *Inf Vis* 4(2):136–146
29. Keim D, Andrienko G, Fekete JD, Görg C, Kohlhammer J, Melançon G (2008) Visual analytics: definition, process, and challenges. In: Kerren A, Stasko J, Fekete JD, North C (eds) *Information visualization, lecture notes in computer science*, vol 4950. Springer, Berlin, pp 154–175
30. Kisilevich S, Krstajic M, Keim D, Andrienko N, Andrienko G (2010) Event-based analysis of people’s activities and behavior using flickr and panoramio geotagged photo collections. In: 2010 14th international conference information visualisation (IV). IEEE, pp 289–296
31. Knox EG, Bartlett MS (1964) The detection of space-time interactions. *Appl Stat* 13:25–30
32. Kraak MJ, Ormeling F (2003) *Cartography: visualisation of geospatial data*. Essex: Pearson Education Limited
33. Lahouari K, Jean-Yves B, Paule-Annick D, Hélène M, Cécile SM (2014) Représenter les dynamiques des territoires : un état des lieux, de nouveaux enjeux. <http://www.map.cnrs.fr/jyb/puca/>
34. Leipnik MR, Albert DP (2003) *GIS in law enforcement: implementation issues and case studies*. CRC Press, London
35. Li H, Zhang J, Sun J (2016a) A visual analytics approach for deterioration risk analysis of ancient frescoes. *J Vis* 19(3):529–542
36. Li M, Bao Z, Sellis T, Yan S, Zhang R (2018) Homeseeker: a visual analytics system of real estate data. *J Vis Lang Comput* 45:1–16
37. Li S, Dragicevic S, Castro FA, Sester M, Winter S, Coltekin A, Pettit C, Jiang B, Haworth J, Stein A et al (2016b) Geospatial big data handling theory and methods: a review and research challenges. *ISPRS J Photogramm Remote Sens* 115:119–133
38. Lins L, Klosowski JT, Scheidegger C (2013) Nanocubes for real-time exploration of spatiotemporal datasets. *IEEE Trans Vis Comput Graph* 19(12):2456–2465
39. MacEachren AM, Jaiswal A, Robinson AC, Pezanowski S, Savelyev A, Mitra P, Zhang X, Blanford J (2011) Senseplace2: Geotwitter analytics support for situational awareness. In: 2011 IEEE conference on visual analytics science and technology (VAST). IEEE, pp 181–190
40. Maciejewski R, Rudolph S, Hafen R, Abusalah A, Yakout M, Ouzzani M, Cleveland WS, Grannis SJ, Ebert DS (2010) A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Trans Vis Comput Graph* 16(2):205–220
41. Malik A, Maciejewski R, Collins TF, Ebert DS (2010) Visual analytics law enforcement toolkit. In: 2010 IEEE international conference on technologies for homeland security (HST). IEEE, pp 222–228
42. Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27(2 Part 1):209–220
43. Miller HJ, Han J (2009) *Geographic data mining and knowledge discovery*. Chapman & Hall/CRC data mining and knowledge discovery series. CRC Press, London
44. Møller J, Ghorbani M (2010) Second-order analysis of structured inhomogeneous spatio-temporal point processes. Tech rep., Department of Mathematical Sciences, Aalborg University
45. Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* 37:17–23
46. Nelson JK, Brewer CA (2017) Evaluating data stability in aggregation structures across spatial scales: revisiting the modifiable areal unit problem. *Cartogr Geogr Inf Sci* 44(1):35–50

47. Openshaw S (1984) The modifiable areal unit problem. *Concepts and techniques in modern geography*
48. Ostfeld RS, Glass GE, Keesing F (2005) Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends Ecol Evol* 20(6):328–336
49. Robinson AC, Peuquet DJ, Pezanowski S, Hardisty FA, Swedberg B (2016) Design and evaluation of a geovisual analytics system for uncovering patterns in spatio-temporal event data. *Cartogr Geogr Inf Sci* 44:1–13
50. Roddick JF, Spiliopoulou M (1999) A bibliography of temporal, spatial and spatio-temporal data mining research. *ACM SIGKDD Explorations Newsletter* 1(1):34–38
51. Scherr M (2008) Multiple and coordinated views in information visualization. *Trends in Information Visualization* 38:1–33
52. Shanbhag P, Rheingans P et al (2005) Temporal visualization of planning polygons for efficient partitioning of geo-spatial data. In: 2005 IEEE symposium on information visualization, 2005. INFOVIS, IEEE, pp 211–218
53. Shekhar S, Jiang Z, Ali RY, Eftelioglu E, Tang X, Gunturi V, Zhou X (2015) Spatiotemporal data mining: a computational perspective. *ISPRS Int J Geo Inf* 4(4):2306–2338
54. Silva R, Moura-Pires J, Santos MY (2012) Spatial clustering in SOLAP systems to enhance map visualization. *Int J Data Warehouse Min* 8(2):23–43
55. Silva R, Pires JM, Santos MY, Datia N (2016) Enhancing exploratory analysis by summarizing spatiotemporal events across multiple levels of detail. In: Sarjakoski T, Santos MY, Sarjakoski TL (eds) *Geospatial data in a changing world, selected papers of the 19th AGILE conference on geographic information science, Lecture Notes in Geoinformation and Cartography*, Springer. https://doi.org/10.1007/978-3-319-33783-8_13, https://link.springer.com/chapter/10.1007/978-3-319-33783-8_13
56. Silva RA, Pires JM, Santos MY (2015a) A granularity theory for modelling spatio-temporal phenomena at multiple levels of detail. *International Journal of Business Intelligence and Data Mining* 10(1):33
57. Silva RA, Pires JM, Santos MY, Leal R (2015b) Aggregating spatio-temporal phenomena at multiple levels of detail. In: *AGILE 2015, Springer Science +mathplus business media*, pp 291–308
58. Sips M, Köthür P, Unger A, Hege HC, Dransch D (2012) A visual analytics approach to multiscale exploration of environmental time series. *IEEE Trans Vis Comput Graph* 18(12):2899–2907
59. Swedberg B, Peuquet D (2016) Perse visual analytics for calendar related spatiotemporal periodicity detection and analysis. *GeoInformatica* 21:1–21
60. Swedberg B, Peuquet D (2017) An evaluation of a visual analytics prototype for calendar-related spatiotemporal periodicity detection and analysis. *Cartographica: The International Journal for Geographic Information and Geovisualization* 52(1):63–79
61. Thakur S, Rhyne TM (2009) Data vases: 2d and 3d plots for visualizing multiple time series. In: *International symposium on visual computing*. Springer, pp 929–938
62. Thom D, Bosch H, Koch S, Wörner M, Ertl T (2012) Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In: 2012 IEEE Pacific visualization symposium (pacifivis), IEEE, pp 41–48
63. Tominski C, Schulz HJ (2012) The great wall of space-time. In: Goesele M, Grosch T, Theisel H, Toennies K, Preim B (eds) *Vision, modeling and visualization, the Eurographics association*. <https://doi.org/10.2312/PE/VMV/VMV12/199-206>
64. Tominski C, Schulze-Wollgast P, Schumann H (2005) 3d information visualization for time dependent data on maps. In: *Ninth international conference on information visualisation, 2005. Proceedings*. IEEE, pp 175–181
65. Tversky B, Morrison JB, Betrancourt M (2002) Animation: can it facilitate? *Int J Hum Comput Stud* 57(4):247–262
66. Wang D, Ding W, Lo H, Morabito M, Chen P, Salazar J, Stepinski T (2013) Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns. *Comput Environ Urban Syst* 39:93–106
67. Weaver C (2010) Cross-filtered views for multidimensional visual analysis. *IEEE Trans Vis Comput Graph* 16(2):192–204
68. Yao JT, Vasilakos AV, Pedrycz W (2013) Granular computing: perspectives and challenges. *IEEE Transactions on Cybernetics* 43(6):1977–1989
69. Yin J, Gao Y, Du Z, Wang S (2016) Exploring multi-scale spatiotemporal twitter user mobility patterns with a visual-analytics approach. *ISPRS Int J Geo Inf* 5(10):187
70. Zhang L, Stoffel A, Behrisch M, Mittelstadt S, Schreck T, Pomp R, Weber S, Last H, Keim D (2012) Visual analytics for the big data era—a comparative review of state-of-the-art commercial systems. In: 2012 IEEE conference on visual analytics science and technology (VAST). IEEE, pp 173–182

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Ricardo Almeida Silva is an invited Assistant Professor on the computer science department at ISEL in Portugal. He has a Ph.D in computer science from Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa (2017). He is also a member of the research center NOVA LINGS where his research is focused on getting knowledge from spatiotemporal data, which involves modeling, data mining and visualization. He was involved in the GeoInsight Analytics Platform project co-funded by NSRF and, in the EU-funded OPTIMUM project. His research focus on visual analytics area, a multidisciplinary field including knowledge discovery, spatio-temporal models, spatio-temporal knowledge representation, geovisualization, information visualization, among others.



João Moura Pires is (since July 2000) Auxiliary Professor of Computer Science in the Department of Computer Science of the Faculty of Science and Technology of the New University of Lisbon. He got a B.Sc. in Electronic Engineering in 1984 by the University of Angola (and its equivalence by University of Porto). In 1991 got an M.Sc. equivalent by the New University of Lisbon. In 2000 he got a Ph.D. in Computer Science, by the the FCT NOVA University, Lisbon.

His current research interest lies in the following topics: Data analytics, Data Mining and Data Visualization and, in particular, Spatio-Temporal Analytics and visualization. Previous research interests include Information Systems, Business Intelligence, Constraint Programming, Fuzzy sets and Fuzzy Logic.

He successfully supervised 3 PhD students and more than 30 of MCs students. He organized or collaborated in the organisation of several Scientific events.

His promoting an informal research network on the topic of Spatio Temporal Analytics - check <http://staresearch.net>



Nuno Datia started teaching at ISEL back in 2001. Since then, he lectures fundamental topics like Programming and Computer Graphics, but also Internet Programming, Information Systems and Data Mining.

He was a member of training teams that deliver courses to industry partners (e.g. Siemens) and to university staff, mainly about the .NET framework.

He got the PhD at FCT-UNL in Computer Science. Currently, I am an associate member at NOVA Lincs research center, a member of the Spatio Temporal Analytics Research interest group, a member of the Future Internet Technologies interest group, an adjunct professor at ISEL and the coordinator of the Master in Informatics and Computers Engineering at ISEL.

Research Interest: Data mining and knowledge discovery, Data warehouse, Data Analytics, Spatiotemporal analysis, Visualisation



Maribel Yasmina Santos is an Associate Professor at the Department of Information Systems, University of Minho, Portugal, where she is also a Senior Researcher at the ALGORITMI Research Centre (<http://algoritmi.uminho.pt>) and at the Engineering Process Maturity and Quality (EPMQ) domain of the Computer Graphics Centre (<http://www.ccg.pt>). She was the Secretary-General of the Association of Geographic Information Laboratories for Europe (www.agile-online.org) from May 2013 to June 2015 and Associate Director of the Department of Information Systems, University of Minho, from May 2010 to July 2014. Currently, she is the leader of the Softwarebased Information Systems Engineering and Management Group at the ALGORITMI Research Centre. Her research interests include Business Intelligence and Analytics, Big Data Analytics, (Big) Data Warehousing, and (Spatiotemporal) Online Analytical Processing. She participated in more than 20 research projects working with data systems and ways to enhance data analytics for decision support.



Bruno Martins is an assistant professor at the Computer Science and Engineering Department of IST and a researcher at the Information and Decision Support Systems Lab of INESC-ID, where he works on problems related to the general areas of information retrieval, text mining, and the geographical information sciences. He has been involved in several research projects related to geospatial aspects in information access and retrieval, and he has accumulated a significant expertise in addressing challenges at the intersection of information retrieval and the geographical information sciences (i.e., in an area that is often referred to as geographical information retrieval).



Fernando Birra is an Assistant Professor working mainly in the field of Computer Graphics. He works in the Department of Informatics of the Faculty of Sciences and Technology of the New University of Lisbon. For a taste of Portuguese, it is called Departamento de Informática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa. He is also a Researcher at NOVA LINCS. His research is conducted in Computer Graphics and Interfaces, with sporadic digressions in Parallel Computing (with GPUs of course). He works with the Multimodal Systems group of NOVA LINCS's centre.