

Special Section on SIBGRAPI 2017

Visual analytics of time-varying multivariate ionospheric scintillation data



Aurea Soriano-Vargas^{a,*}, Bruno C. Vani^b, Milton H. Shimabukuro^b, João F. G. Monico^b,
 Maria Cristina F. Oliveira^a, Bernd Hamann^c

^a Instituto de Ciências Matemáticas e de Computação (ICMC), University of São Paulo (USP), São Carlos, SP, 13566-590, Brazil

^b Faculdade de Ciências e Tecnologia (FCT), São Paulo State University (UNESP), Presidente Prudente, SP, 19060-900, Brazil

^c Department of Computer Science, University of California, Davis, CA 95616, USA

ARTICLE INFO

Article history:

Received 26 March 2017

Revised 15 August 2017

Accepted 15 August 2017

Available online 24 August 2017

Keywords:

Visual analytics

Time-varying multivariate data

Visual feature selection

Data visualization

Ionospheric scintillation

ABSTRACT

We present a clustering-based interactive approach to multivariate data analysis, motivated by the specific needs of scintillation data. Ionospheric scintillation is a rapid variation in the amplitude and/or phase of radio signals traveling through the ionosphere. This spatial and time-varying phenomenon is of great interest since it affects the reception quality of satellite signals. Specialized receivers at strategic regions can track multiple variables related to this phenomenon, generating a database of observations of regional ionospheric scintillation. We introduce a visual analytics solution to support analysis of such data, keeping in mind the general applicability of our approach to similar multivariate data analysis situations.

Taking into account typical user questions, we combine visualization and data mining algorithms that satisfy these goals: (i) derive a representation of the variables monitored that conveys their behavior in detail, at multiple user-defined aggregation levels; (ii) provide overviews of multiple variables regarding their behavioral similarity over selected time periods; (iii) support users when identifying representative variables for characterizing scintillation behavior. We illustrate the capabilities of our proposed framework by presenting case studies driven directly by questions formulated by collaborating domain experts.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Ionospheric scintillation is an atmospheric phenomenon that affects measurements obtained by Global Navigation Satellite Systems (GNSS) such as the Global Positioning System (GPS, U.S.A.), Global Navigation Satellite System (GLONASS, Russia), Galileo (European Union) and BeiDou Navigation Satellite System (BDS, China). It results from amplitude attenuation and phase shifts in GNSS radio signals as they propagate through regions of the ionosphere with irregular electron densities [1]. When scintillation occurs GNSS receivers may experience a complete *loss of lock* on the affected signals, i.e., there can be a discontinuity in the phase tracking loop [2], as well as a degradation in the accuracy of the measurements from a given satellite [3]. Such effects preclude satellite availability for positioning purposes, degrade the positional accuracy and may lead to service outages at critical circumstances, losing track of one or more satellites. In critical circumstances, the

positioning service can be lost when a minimum number of satellites are not successfully tracked by the receiver [4].

Latin America, and in particular regions in Brazil located around the magnetic equator are severely affected by a frequent and strong amplitude scintillation [5,6]. Consequently, applications that rely on GNSS technology and require full availability and accuracy for their operation – such as Precise Point Positioning (PPP), Real Time Kinematic (RTK) applied to land surveying in agriculture and differential GNSS applied to offshore oil surveying – may face significant and potentially damaging issues. Analyzing the occurrence patterns of scintillation and understanding their causes is essential in order to gather elements for addressing its effects.

A network of Ionospheric Scintillation Monitor Receivers (ISMRs) has been operating in Brazil since 2011 to monitor the phenomenon [2], supported by projects CIGALA (Concept for Ionospheric Scintillation Mitigation for Professional GNSS in Latin America) and CALIBRA (Countering GNSS high Accuracy applications Limitations due to Ionospheric disturbances in Brazil)¹.

* Corresponding author.

E-mail addresses: asoriano@icmc.usp.br, aurea.soriano@gmail.com

(A. Soriano-Vargas), brunovani@ifsp.edu.br (B.C. Vani), miltonhs@fct.unesp.br (M.H. Shimabukuro), galera@fct.unesp.br (J.F. G. Monico), cristina@icmc.usp.br (M.C. F. Oliveira), hamann@cs.ucdavis.edu (B. Hamann).

¹ Both projects have been funded by the European Commission under the framework of the FP7-GALILEO-2009-GSA and FP7-GALILEO-2011-GSA-1a, respectively.

The ISMRs include GNSS capabilities specially designed to provide scintillation metrics. A typical receiver - PolaRxS PRO, manufactured by Septentrio - monitors 62 variables computed every minute from data sampled at high rates (50 Hz), plus eight additional variables are calculated every minute. Thus, 70 variables related to scintillation are continuously measured and computed every minute for all 187 satellites tracked, which include 32 GPS satellites. Data collection occurs at 12 monitoring stations spatially distributed in Brazil. The resulting database records an expressive volume of historical observations of the temporal behavior of scintillation indices and related variables, where each observation consists of measurements of multiple variables at a particular time.

This data, which we refer to as the ISMRs database, has been available to experts for some time and visualization tools have already been developed to support data analysis [4]. Nonetheless, domain experts still lack additional support for exploratory data analysis to study the interplay between the multiple variables and their role in characterizing scintillation behavior [7]. We introduce a visual analytics approach to support exploration of historical ionospheric scintillation data. It integrates multiple visualization and data mining techniques into a visual exploration loop that allows scintillation scientists to gain an understanding of how the multiple variables tracked are related over short and extended time periods, and how they relate to observed behavior of scintillation measures.

The driving requirements guiding the design and development of our solution approach were: (i) To define a representation of the variables monitored that conveys their behavior in detail, at multiple user-defined aggregation levels; (ii) to provide overviews of multiple variables regarding their behavioral similarity over selected time periods; (iii) to support users when identifying representative variables for characterizing scintillation behavior. These requirements have been addressed by integrating various visualization techniques and mining algorithms.

A so-called time matrix visualization shows the behavior of a particular variable over a time period, while still allowing a user to inspect specific individual values. The visualization can be created directly from the recorded values or from values aggregated for user-defined temporal units, to support observation of extended time periods. A small multiples time matrix visualization simultaneously depicting a subset of user-selected target variables is also shown, where the individual matrix views are spatially grouped to highlight the similar/dissimilar temporal behaviors of groups of variables. A similarity map visualization of the variables is provided as well, complementing and summarizing the small multiples view. These multiple views are coordinated and analysts can explore them jointly to identify representative subsets of variables to characterize scintillation behavior, and assess feature subspaces in relation to scintillation measures by means of classification algorithms.

We contribute new strategies for visualizing time-varying multivariate data sets. These strategies allow experts studying the ionospheric scintillation phenomenon to utilize alternative approaches for exploring a large database of historical observations. Our approaches complement existing ones [4] by focusing on the global temporal relationships between the multiple variables tracked, rather than on their individual behavior in isolation.

The paper is structured as follows: Section 2 discusses related work on visual analytics applied to feature selection problems and in visual analytics of time-varying multivariate data. We also review previous contributions that addressed analysis of collected scintillation data. In Section 3 we describe in details the ionospheric scintillation data and the preprocessing steps. Our visual analytics approach, comprising feature extraction, clustering, individual visualization and global visualization of data variables is de-

scribed in Section 5. In Section 6 we illustrate possible applications of the proposed visual exploration solution to plausible data analysis scenarios and assess its capability to inform the relevant variables for explaining ionospheric scintillation. Conclusions and a discussion are provided in Section 7.

2. Related work

Previous research efforts related to the topics addressed in this paper are found in the fields of feature selection assisted by visualization and visualization of multivariate time-varying data. Also relevant is previous work in mining and visualization of ionospheric scintillation data.

2.1. Visual feature selection

Algorithms for analysis and visualization of multidimensional data typically face the problem known as the *curse of dimensionality* [8], which hampers a clear interpretation of the role of individual variables (data features) and their interaction in producing the data patterns. High-dimensional data is likely to be described by redundant variables and feature selection algorithms are widely employed for dimension reduction. The challenge is to identify a reduced subset of features sufficient to describe the intrinsic data space and determine data behavior [9]. The best possible subset would include a minimum number of features that contribute mostly to accuracy of a classifier or a regressor [10].

Razente et al. [11] advocate that integrating dimension reduction techniques with visualization strategies offers great potential to support analysts in tasks of identifying representative data attributes. Several research contributions couple interactive graphical representations with feature selection processes in favor of improved understanding of data behavior.

Some authors approach visual feature selection relying on statistical techniques and distances combined with visualizations [12–15]. Whereas some contributions are domain specific, e.g., the VIDEAN system [15] integrates several coordinated visual representations to assist feature selection in a cheminformatics problem, others provide general-purpose visual interfaces for feature subset selection. One example is SmartStripes [14], developed for diagnostic purposes, providing interactive refinement of automatic feature subset selection techniques, considering the interplay between different feature and entity subsets. Some systems [12,13] implement ranking-based visual strategies to identify similarities among variables.

Other contributions [16–18] concern the problem of analyzing and comparing different variable subspaces, and subspace clustering for the analysis of high-dimensional data, as reviewed by Liu et al. [19]. Approaches such as representative factor generation [17] and dimension projection matrix/tree [18] allow interactive exploration of both data variables and data observations in order to investigate how variables are related. Typically, time-varying multivariate features are not explicitly handled.

Our approach is also a contribution in visual feature selection and analysis of feature subspaces, but it incorporates the dynamics of temporal variations into the process. As scintillation is a seasonal phenomenon, the role of different variables in characterizing its behavior changes along time. We contribute a solution for visual feature subspace analysis in time-varying data sets, which also gives analysts the capability of including their domain knowledge into the investigation.

2.2. Visualization of time-varying multivariate data

Multivariate time-varying data plays an important role in many application domains and poses multiple challenges to visualiza-

tion designers [20]. The continuous collection of multiple variables typically results in bulky data sets that exhibit complex behavior, making the design of visual interfaces to guide specialists in search of relevant data patterns particularly challenging.

Several classical multivariate visualization techniques have been employed to convey time-related information, e.g., Parallel Coordinates [21] used in connection with an algorithm to identify important temporal trends in multivariate scientific data sets [22], or combined with histograms to quantify visual properties and reduce visual overload in time-varying volume data [23].

Small multiple visualizations rely on the repetition of the same design structure to summarize and compare variables [24]. Correlated multiples [25] is a method that adopts a spatially coherent placement of the multiples views, where their relative distances reflect their dissimilarities. It has been applied to univariate data from three different domains, namely stock market trends, census demographics and climate modeling, in order to assess changes over the years and identify common trends and similar years. TimeSpiral [26] is a visualization system combining multiple views to assist users in analyzing and exploring periodic trends and correlations in multivariate time-series data. It supports data aggregation at multiple levels, e.g., changing a time interval, time span or time granularity. These systems are suitable for simultaneous investigation of a reduced number of variables.

VIMTEX has been designed to assist geologists observing temporal relationships in multivariate data describing concentrations of chemical compounds [27]. It uses Parallel Coordinates for a multivariate, time-varying view of the data, combined with a density view to show univariate temporal distribution and a small multiples matrix view which shows bivariate correlations as time-series. The interactive system Falcon [28] coordinates time-oriented and statistical views for users to explore temporal and statistical patterns in multiple time-varying variables associated with 3D printing processes.

Several systems have been introduced for climate and climate modeling data analysis. Vismate is a visual analytics system for exploring climate changes in P.R. China at different spatio-temporal scales [29]. It uses land surface observations collected by meteorological observation stations, and combines three visualizations: (i) A Global Radial Map using K-means clustering; (ii) a Time-Series Discs using multiple time-series and triangular HeatMaps around a center point; and (iii) an Anomaly Detection Scatterplot based on Principal Components Analysis. The Similarity Explorer [30] combines small multiples visualizations and coordinated views for visual comparison of the outputs resulting from simulations of multiple climate models. The tool supports spatio-temporal exploration focusing on the analysis of correlations between climate models with respect to any variable.

We also present a domain-specific solution for scintillation data that includes several strategies available in existing systems. For example, similarly to known solutions [25,30] we adopt a small multiples visualization and a consistent placement strategy to display groups of related variables; and we consider arbitrary time periods of different granularity levels by means of user-defined data aggregations [26]. Our solution, however, supports the simultaneous investigation of many variables and satisfies specific requirements of scintillation experts. It provides the ability to inspect and explore scintillation data aggregated over different temporal scales, to investigate the behavior of variables individually or in groups, and to explore alternative feature spaces for characterizing the scintillation phenomenon. Although the introduced system is domain-specific, it supports tasks that are also applicable to the analysis of other multivariate time-varying data in other domains, as discussed in Section 7.

Table 1
Categorization of S4 index intensity by Tiwari (2011) [32].

S4 values	Scintillation Categorization
$S4 \geq 1.0$	High
$0.5 < S4 < 1.0$	Moderate
$S4 \leq 0.5$	Low

2.3. Analysis of ionospheric scintillation data

The so-called S4 index of scintillation amplitude is often used to measure the intensity of ionospheric scintillation. It is computed as the standard deviation of the signal intensity normalized by its mean [31]. Signal intensity must be measured at high rates for the index to detect rapid fluctuations. Currently, there is no consistency concerning the categorization of the severity of scintillation as measured by the S4 index, Table 1 describes one well-accepted categorization.

Several authors applied data mining techniques to collected scintillation data, e.g., Rezende et al. [7] devised a method that combines a bagging method, which uses bootstrap to randomly generate several samples from an original sample, with decision trees to predict the S4 index with one or more hours of antecedence. Lima et al. [33] presented a correlation analysis of occurrences of ionospheric scintillation registered in two stations at different locations in Brazil. They used a classification and regression decision tree (CART) from the S4 index. Another recently published technique [34] uses neural networks to predict two levels of scintillation: strong or not strong (low/moderate). Analyses are typically supported by simple univariate time series visualizations.

Ackah et al. [35] investigated records of the S4 and vertical TEC (VTEC) indices in a West African equatorial region. The presented method visually represents the time series as grids where a column represents a daily hour, a row represents a particular day, and grid cell color maps an index (S4 or VTEC) at the corresponding day/hour. We have adopted similar grid representations as well to convey variable values measured over a time range.

The ISMR Query Tool [4,36] is an integrated software platform specifically developed for the ISMRs database. It includes four visualizations of the S4 index: a scatter-plot view of data from one or multiple satellites over a time period, a calendar view, a Ionospheric Pierce Point (IPP) representation, aggregated time series visualizations obtained with the SAX approach; and a horizon chart visualization of one or multiple variables. However, it does not focus on multivariate analysis or on the interplay of the many variables characterizing the scintillation phenomenon. Our solution complements that previous effort by providing strategies to inspect and compare the historical behavior of multiple variable subspaces in relation to several ionospheric indices, including, but not limited to, the S4 index.

3. Data description and processing

Scintillation observations recorded in the ISMRs database result from monitoring 187 GNSS satellites tracked by receivers placed in 12 monitoring stations distributed in Brazil, indicated in the map in Fig. 1. As mentioned in Section 1, observations recorded in the data base consist of minute-by-minute measurements (or computations) of 70 variables, for each satellite tracked. The set of observations can be interpreted as multivariate time series of numerical values describing multiple scintillation indices (46 variables), measures of the S4 index (12 variables), measures of the SigmaPhi index (three variables), time stamps (seven variables), and indicators of a satellite's spatial orientation (two variables).



Fig. 1. Locations of CIGALA and CALIBRA stations in Brazil. PRU1 and PRU2 stations are located in the State of São Paulo.

The scintillation indices describe standard deviation of phase computed over different time intervals; average and standard deviation of code/carrier divergence; spectral strength and spectral slope of detrended phase; averaged signal-to-noise ratio; and absolute and differential total electron content. The S4 indices measure amplitude scintillation in different bands, as the standard deviation of the detrended signal intensity (total S4), the thermal noise correction of the amplitude scintillation (S4 correction), and the corrected indices (corrected S4). The SigmaPhi indices measure phase scintillation in different bands as the standard deviation of the detrended carrier phase from received GPS signals. The time stamps group consists of the lock time (synchronization time) between satellites and receiver in different bands and GPS time. The direction indicator group refers to satellite azimuth and elevation angles relative to the receiver. When retrieving observations it is convenient to apply an elevation cut-off to disregard data from lower satellites, which are more susceptible to noise or multi-path effects [37].

We can formally describe the scintillation data set as an instance of a time-varying data set with n variables $\{x_1, x_2, x_3, \dots, x_n\}$, each described by a sequence of time-stamped numerical values, i.e.,

$$x_i = \{x_i^{t_1}, x_i^{t_2}, x_i^{t_3}, \dots, x_i^{t_k}\}. \quad (1)$$

A data observation $\mathbf{o}^{(t)}$ is defined as a vector consisting of the values of p selected variables, $p \leq n$, taken at a time t , i.e.,

$$\mathbf{o}^{(t)} = [x_1^t, x_2^t, x_3^t, \dots, x_p^t]. \quad (2)$$

We also define aggregated data observations $\mathbf{o}^{(t_{se})}$ as a vector of p values, i.e.,

$$\mathbf{o}^{(t_{se})} = [x_1^{t_{se}}, x_2^{t_{se}}, \dots, x_p^{t_{se}}], \quad (3)$$

but now each $x_i^{t_{se}}$ is a value resulting from an aggregation function f applied to the values of x_i in a time interval $[t_s, t_e]$, i.e.,

$$x_i^{t_{se}} = f(x_i^{t_s}, x_i^{t_s+1}, x_i^{t_s+2}, \dots, x_i^{t_e}). \quad (4)$$

Many aggregation functions are possible, e.g., maximum, minimum, average, median or standard deviation. From the previous definitions, scintillation observations (aggregated or not) can be described by distinct variable subspaces, according to the interests of the analyst.

Since variables have values on different dynamic ranges, with means and variances on different orders of magnitude, data ranges are linearly normalized to the interval $[0,1]$ prior to performing any further processing.

Table 2

High-level analysis questions identified in collaboration with domain experts.

Q1	Which variables tracked show similar behavior over a target time period?
Q2	How do variables relate to each other and to the scintillation indices?
Q3	Besides the S4, what (minimal) subset of variables suffices to characterize scintillation?
Q4	Do the same variables tracked by different satellites show similar behavior?

The case studies presented in Section 6 consider data recorded at two stations, PRU1 and PRU2, located at Presidente Prudente, in the State of SˆAo Paulo, see Fig. 1. For practical reasons, we downloaded data from the ISMRs database to a local relational database. So far we have explored data from the 32 satellites in the GPS constellation, which is currently the world's most widely employed satellite navigation system [38]. Nonetheless, the solution we describe is applicable to observations provided by any satellite or monitoring station.

Missing values are quite common and need special treatment. In some cases, they characterize satellite behavior, due to satellites having different orbital periods, and sometimes they are due to reception errors. We consider one of three strategies for handling the missing-value problem: (i) If a value at a particular time point is missing, but values are known for its previous and subsequent time points, its value will be linearly interpolated from the neighboring values; (ii) if values for a whole day are missing, but values are known for its previous and subsequent days, values for the day will be linearly interpolated from the corresponding neighboring values from the previous and subsequent days, accounting for the time shift; (iii) if the previous cases do not apply, values will not be estimated and the respective entries remain unknown.

4. Driving questions and goals

In partnership with domain expert collaborators we identified prototypical data analysis questions that cannot be answered with existing tools. These questions, listed in Table 2, suggest that analysis should be able to inspect the temporal behavior of both individual variables and of groups of related variables, as well as investigate which variables are most relevant to characterize the behavior of the diverse scintillation indices measured.

The ability to provide answers to these questions defined the requirements for visualization design. Our collaborators were interested not only in inspecting a global overview of variables' behavior, but wanted to be able to identify specific values and time points along the exploration. Thus, it was important to capture the multivariate nature and temporal behavior without losing detail, e.g., being able to probe specific values. Keeping this in mind, each matrix visualization uses enough display space to convey the details over time, whereas the small multiples view provides the overview and ability to compare variables in terms of their behavior. The similarity maps are also a proper choice to convey a global view of the relationships between variables in terms of their temporal behavior. Therefore, our design efforts concentrated on providing complementary views of multivariate relationships and allowing users to explore representative variable sub-spaces in relation to scintillation behavior. In Section 6, we demonstrate the system's resources by means of three exploration cases conducted to exemplify how it can be helpful to finding answers to such questions.

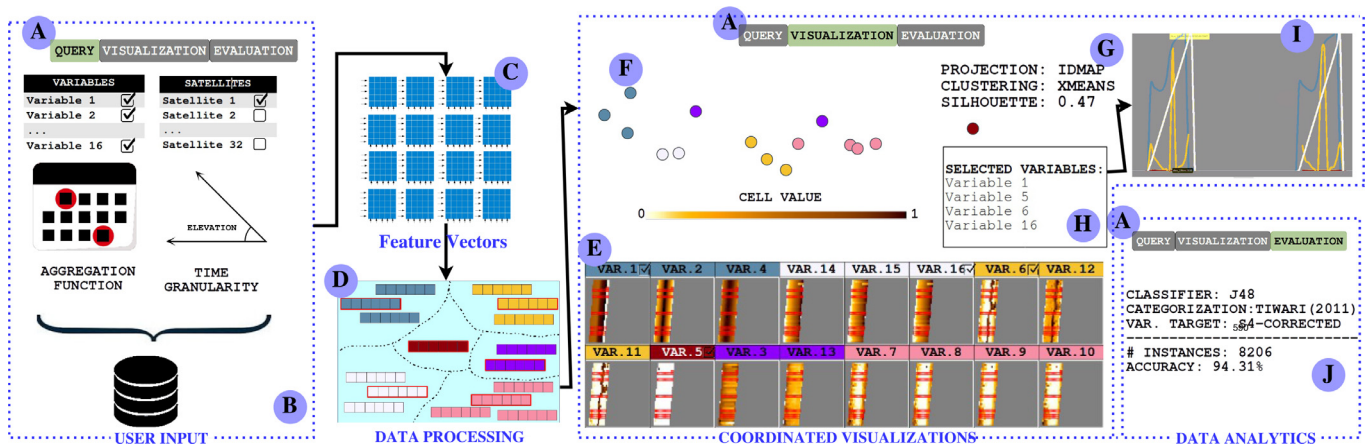


Fig. 2. Overview of the data processing and interface functionalities (listed in the menu, labeled A). Initially, a query panel (B) allows one to retrieve a subset of observations for a time range, described by a subset of variables (sixteen being selected in the example). The regions labeled C and D describe the underlying data processing steps: values are normalized and a time matrix representation is derived for each variable (C). Each time matrix is described as a multidimensional feature vector, and feature vectors are clustered into groups of variables that exhibit similar behavior over the observed time range (D). Multiple coordinated visualizations can be explored: region E shows a small multiples time matrix view of all variables, whereas region F shows a similarity map (projection) view, with variables represented as circles. The color mapping in the matrix headers and circles reflect the variables' assigned cluster. Region G shows functions to modify the cluster model or select other projection techniques. Standard timeline views of four user-selected variables are shown that reveal seasonal patterns (H, I), and classification and regression functions are also available (J).

5. Visual analytics solution

We have integrated mining and visualization techniques into a framework to support experts in the investigation of the temporal behavior of data observations, as defined by Eq. (2) or (3) in Section 3. Observations described by alternative variable subspaces can be inspected focusing on the time-varying behavior of either individual variables or groups of variables. This flexibility is a key strength in supporting extensive data exploration tasks considering multiple time periods and temporal scales, as well as alternative variable subspaces for characterizing the phenomenon.

The introduced system combines a small multiples visualization component that conveys the temporal behavior of each variable with clustering and multidimensional projection techniques in a consistent representation of the data observations, in order to convey temporal patterns and highlight representative variables for characterizing the observed scintillation behavior, as measured by the pertinent indices.

The system is provided as a web application developed using Java and Java Server Pages (JSP) on the server-side, PostgreSQL 9.3.9 as database, and HTML5, CSS and Javascript on the client-side. SQL functions with subscripts are employed to retrieve aggregated data values relative to a specified time period.

In the following, we provide an overview of the data exploration functionalities and describe the visualizations and the steps involved in creating them.

5.1. Overview of the visual analytics system

Fig. 2 provides an overview of the steps involved in creating the visualizations and the interface functions for exploring the data. The main interface menu is shown in the region labeled A. Initially, the user issues a query against the database by specifying a target time period for retrieval (initial and final days) and what subset of variables describe the observations (region B). (It is possible to select all variables.) The query must also specify (i) the source satellite(s) from which measurements must be recovered and an elevation angle cutoff; the time span for data aggregation, considering that observations are recorded by minute and may be aggregated over arbitrary time spans (10 minutes, one hour, one day, etc.); and (iii) the aggregation function to summarize the values for the se-

lected time span (e.g., average, median, standard deviation, minimum value, or maximum value).

The regions labeled C and D in the figure refer to the underlying data processing steps. The time matrices are created as temporal representations of each variable, where each entry stores a single (possibly aggregated) normalized value (region C) relative to the defined time span, providing the basis for a detailed view of each individual variable over time and a small multiples visualization of subsets of variables.

The time matrices are represented by multidimensional feature vectors, which are input to a clustering algorithm (region D) to identify groups of variables that show similar behavior over the observed time period. From these representations multiple visualizations are created for data exploration: region E shows a small multiples visualization of the time matrices relative to each variable observed (with header colors indicating their corresponding cluster), whereas region F shows a similarity map view of all variables, created from their multidimensional feature vectors. Each circle represents a variable, colored to indicate its cluster. Both views can be inspected and manipulated to refine the current cluster model to reflect user knowledge and beliefs. Region G shows information regarding the quality of the clustering and functions to modify the cluster model or select other projection techniques for the similarity map. Region H shows that four variables have been user-selected, for which timeline views are displayed (region I) revealing seasonal patterns. It is possible to apply classifiers and regressions to the scintillation indices (J). Iterative data analysis with these multiple coordinated visualizations, combined with results of clustering, classification and regression algorithms, allow experts to assess the role of different variable sub-spaces as representative descriptors of the behavior of the ionospheric scintillation indices.

5.2. Time matrix visualizations

A time matrix representation summarizes the temporal behavior of a particular variable for a user-defined observed period. As the user may choose to aggregate a range of observed values, a time matrix maps the values of a variable along a certain period and at a given time granularity, from minutes to hours or even days. Each matrix entry represents a single numerical value, either as recorded in the database or derived according to Eq. (4), applying an aggregation function over a range of values.

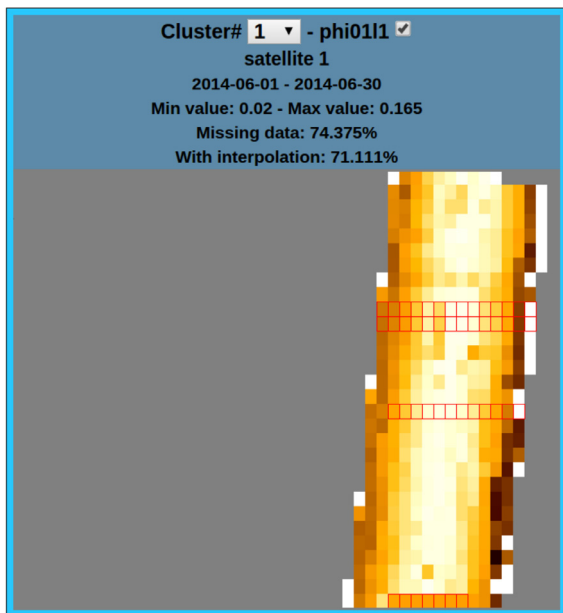


Fig. 3. Time matrix visualization of variable ($\phi 0111$). Daily data (00:00h-23:00h) from GPS satellite 1 at station PRU2, from June 1 to 30, 2014. In this example, cells depict the maximum value observed over a 30-minute interval (user choice of aggregation function and time span). Each row depicts values for a day and each column shows values for the same thirty-minute period, on different days. Areas in gray depict matrix cells with missing values, and cells enclosed by a red border show interpolated data. The matrix border in cyan indicates a high rate of missing data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Each time matrix has a corresponding visualization (Figs. 3 and 4) which is split into two areas. The header area shows the variable's name, its assigned cluster (also indicated by the header and border color), the satellite from which observations have been retrieved, the observation period, the minimum and maximum values recorded over that period and the rate (percentage over total) of missing data before and after interpolation.

The main area is split into cells depicting corresponding matrix entries. Values are color-mapped using the heated-object colormap [39], where brighter colors map to lower values and darker colors map to higher ones (Fig. 5).

Cells that show interpolated data are highlighted by a red border, and cells depicting unknown entries (missing data) are shown in gray without a border. The border of a time matrix may be highlighted in cyan to indicate it has a missing data rate above a user-defined threshold (say, above 50%). Our choice of colors is based on red and cyan being recommended colors to catch viewer attention [40]. They are also complementary hues [39]. Of course, these choices may not be ideal in all situations.

In the examples, cells show the maximum value observed within a 30-minute interval (Fig. 3) and the median value observed over a day (Fig. 4). Hovering over a cell causes a label to show up with detailed information about the cell's corresponding entry: date, recorded value and aggregation function and time span. This is illustrated for the two extreme cells of the time matrix depicted in Fig. 4.

We compute a feature vector from each time matrix, formed by the five standard statistical moments computed from the matrix, i.e., mean, standard deviation, skewness, kurtosis, and uniformity. The statistical moments capture the properties of the probability distributions that characterize the variables' behavior over the observed time period. Matrix entries with missing values are ignored in the computation of statistical moments. However, there is an option for including the percentage of missing data as an ex-

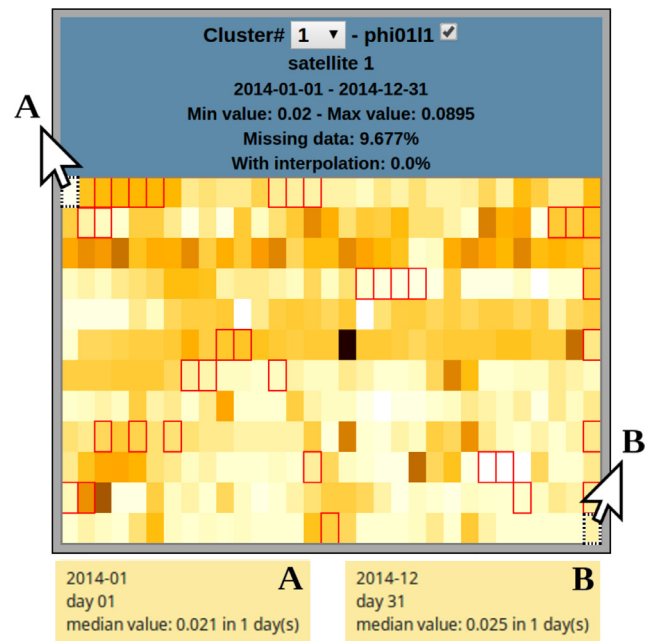


Fig. 4. Another time matrix visualization of the same variable ($\phi 0111$), now depicting monthly data from January 1 to December 31, 2014. In this example, each cell depicts the median value observed over a day. Each row shows daily values observed along a specific month and each column shows values relative to the same sequential day along the 12 months. Regions marked A and B show the detail labels overlaid when the mouse is placed over the indicated cells (shown separately for clarity).



Fig. 5. Heated-Object colormap [39]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tra feature in the matrix feature vector, as in some situations such a variable should be clearly identified as an outlier.

We cluster feature vectors in order to group variables based on global similarity of their observed behavior. We offer a choice between K-means, Bisecting K-means, or X-means clustering algorithms and use the Euclidean distance between feature vectors as a measure of their dissimilarity. The advantage of X-means clustering [41] is the fact that it does not require specifying a desired number of clusters *a priori*, as it learns this number as the one that yields the best Bayesian information criterion. Users may re-assign variables to clusters or create new clusters based on their expert perception of joint similarity, e.g., after inspection of time matrix views and similarity maps of variables, as described in the next Subsection. The alternative cluster models can convey insight to analysts attempting to identify relevant variables and variable subspaces for further investigation with classification or regression algorithms.

Time matrix visualizations shown in Figs. 3 and 4 can be viewed individually or as a small multiples visualization of all queried variables, as shown in Fig. 7B, where views are spatially arranged according to their assigned clusters. A combo box, shown in the header area of each matrix view, allows analysts to change the corresponding variable's cluster assignment, in case they disagree with the automatic assignment. A measure of cluster quality, the silhouette coefficient [42], is informed. Silhouette values are in the range $[-1, +1]$, where higher values indicate a cluster model with higher cohesion and separability. Values are updated whenever the current cluster model changes. Silhouette coefficients provide analysts with an objective measure of cluster quality, which may be helpful when investigating and comparing alternative clus-

ter models and being unsure about a model's suitability. Indeed, the analyst's perception of relevance and similarity of a variable's behavior may be different from the one inferred by automatic clustering. Our approach encourages investigating alternatives, where users' knowledge is more important for defining the assignment of variables to clusters than silhouette measure values.

At this stage, we have generated, possibly with user intervention, clusters of variables that behave similarly for an observation period. For subspace analysis purposes, the variables corresponding to the medoids of clusters are initially assumed to be the ideal set of representatives for describing the corresponding observations. Nonetheless, analysts can modify the default choice of representative variables by designating a different variable as a cluster's representative, done by ticking a checkbox shown in the corresponding time matrix view. For example, the ticked checkbox of variable *phi0111* in Figs. 3 and 4 indicates that it is the cluster representative. This can be achieved either by default, because this variable is the medoid of a cluster, or it is a user's choice. Flexibility in defining the cluster model and picking representative variables is important in situations where analysts want to compare alternative feature subspaces to characterize scintillation behavior well. Our system supports a similarity map visualization of variables, described next.

5.3. Similarity map of variables

Similarity map visualizations of the variables can help specialists assessing, and possibly modifying, the clustering results according to their appreciation of similarity in variable behavior. A similarity map depicts a two-dimensional (2D) space embedding of the m -dimensional ($m = 5$ or $m = 6$) feature vectors describing the variables' time matrices, in which 2D point distances are used as proxies to the point distances in the original 6-dimensional space [43,44]. Therefore, "more similar" variables (as described by their corresponding time matrix) are placed closer in the 2D map, whereas more "dissimilar" ones are placed farther apart.

The 2D embedding is computed with the IDMAP multidimensional projection [44] (we used an implementation by the authors). IDMAP combines two techniques: (i) it generates an initial mapping with Fastmap [45], and (ii) iteratively improves point placements using the Force Scheme strategy [43], recovering information lost in the initial projection. Pairwise Euclidean distances are used to define the dissimilarities between feature vectors. Similarly to time matrix visualization, circles can be highlighted with a cyan border to indicate missing data above a threshold.

One such similarity map is illustrated in Fig. 7A. Each colored circle depicts a variable and its associated cluster, preserving the same color coding of the clusters adopted in the small multiples matrix view. Circle borders in cyan map variables with over 50% missing data prior to interpolation.

The map conveys the overall similarity between variables and their implicit groupings, regardless of their explicit cluster assignments. Therefore, it helps analysts in identifying the "natural" group neighborhoods. This is the reason for showing the map view in connection with the small multiples matrix view: inspecting both in combination provides the analyst with insights needed to decide whether the explicit cluster assignments should be modified and alternative cluster models should be investigated.

Several projection techniques can be employed to obtain the similarity maps. Our choice of IDMAP was based on performing a comparative analysis of maps obtained with different multidimensional projection techniques. We employed Neighborhood Preservation (NP) [46] curves to compare techniques regarding their precision in preserving multidimensional neighborhoods. One such comparison is illustrated in Fig. 6. For varying neighborhood sizes k , the NP curve is computed as follows: The k -nearest neighbor

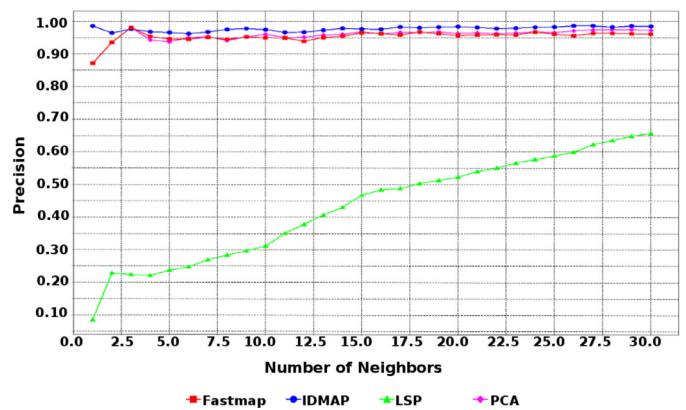


Fig. 6. Neighborhood Preservation curves depicting precision of similarity maps of the 70 variables obtained with four distinct techniques.

sets of each data point in multidimensional and projected space are obtained, and their overlapping rate is computed, yielding values in the interval [0, 1]. The average precision values for all data points for the considered range of k are plotted. The resulting curves will approach maximum precision as k increases, but some techniques can achieve high precision even for smaller values of k .

Fig. 6 shows the NP curves of similarity maps computed with IDMAP and three alternative projection techniques, namely Fastmap [45], Least Squares Projection [47] and Principal Component Analysis [48], for the data investigated in the first case study discussed in Section 6. Clearly, IDMAP yields precision rates superior to LSP and comparable to those obtained by classic approaches such as Fastmap and PCA. Our experiments have indicated that IDMAP usually preserves neighborhood well for this data.

6. Use cases and results

Experts can use the proposed framework to interactively explore recorded scintillation data over different temporal ranges and scales, from multiple perspectives. We limit our discussion to the selected motivating examples raised by the questions listed in Table 2.

6.1. Identifying variables with similar temporal behavior (Q1)

Data for this study was obtained by sampling the maximum values recorded at the station (PRU1), over one-hour periods each day in January 2014 (00:00 h–11:00 h), with an elevation cutoff of 30 degrees, yielding 354,727 observations described by 70 variables. We considered data from satellite GPS 1. The X-means clustering of all 70 variables generated six clusters, with a silhouette coefficient of 0.478, which indicates a high-quality clustering result. The composition of the clusters may be observed inspecting both the similarity map of variables and the small multiples time matrices views, shown in Fig. 7, in which the cluster assignment is indicated by the color of circles or time matrices, respectively.

While the map provides an overview of the variables' global similarities, the small multiples view details their temporal behavior, making it possible to verify the actual behavior of cluster members and assess their (dis)similarities. Cluster C1 (light blue) contains 12 variables, and it is possible to observe two sub-groups in the map: a cohesive group of nine variables and another sub-group of three variables t_{11} , t_{12c_e5a} and t_{e5b} not very close to each other, but their time matrices showing quite similar behavior. In the pink cluster (C2, with 7 variables), variables $sigmaccd_{11}$, $sigmaccd_{12c_e5a}$ and $sigmaccd_{15_e5b}$ show similar behavior, slightly different from other cluster members. The time ma-

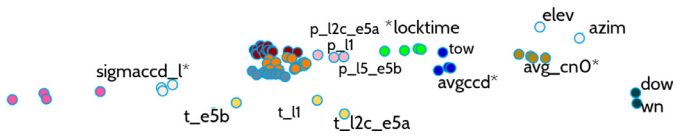


Fig. 8. IDMAP similarity map view of the same 70 variables depicted in Fig. 7A, in which circle colors reflect an alternative (user-adjusted) cluster model (12 clusters). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

trices of the 16 variables in the red cluster (C3) show remarkably similar behavior. In cluster C4 (orange), which groups 19 variables, we notice that p_{l1} , p_{l2c_e5a} and p_{l5_e5b} behave differently from the others. This can be also observed in the similarity map, where the circles representing these three variables are slightly apart from the main group of orange circles. Cluster C5 (dark green) groups two variables related to time (day of week and week number), whereas the variables *azim* and *elevation*, which indicate spatial orientation of satellites, are grouped together into cluster C6 (light rose). The neon green cluster (C7, with nine variables), observed in the map, is also split into two subgroups: one formed by variables avg_cn0_{l1} , avg_cn0_{l2} , $avg_cn0_{l5_e5b}$ and $f2nd_tec_cn0$; another formed by variables $l1_locktime$, $l2_e5a_locktime$ and $l5_e5b_locktime$. When observing their corresponding time matrices one notices that these latter variables show a distinct behavior from the others in this cluster. Also distinctive in this group is the time variable *tow*, placed in the map somewhere in between this subgroup and the dark blue cluster C8. The time matrices of variables in cluster C8 show that they behave similarly, and they are indeed placed close in the map.

These two visualizations when viewed in combination provide useful information for the analyst to adjust the cluster model obtained with X-means. The variables' cluster assignment can be modified using the combo box provided in its time matrix view. The ability to identify clusters of variables with similar behavior is important when investigating answers to the questions discussed next.

Considering the remarks above, the cluster model was modified as follows: cluster C1 (light blue) was split into two subgroups formed by the cohesive group of nine variables and by the three disperse variables, respectively; cluster C2 (pink) was also split into two groups (formed by variables $sigmaccd_{l1}$, $sigmaccd_{l2c_e5a}$ and $sigmaccd_{l5_e5b}$ in a group; and the remaining variables in the other); cluster C4 (orange) was split into two groups (one sub-group with variables p_{l1} , p_{l2c_e5a} and p_{l5_e5b} and another sub-group with the remaining variables); cluster C7 was split into two subgroups (one formed by variables avg_cn0_{l1} , avg_cn0_{l2} , $avg_cn0_{l5_e5b}$ and $f2nd_tec_cn0$ and the other by the remaining variables); and variable *tow* was re-assigned to cluster C8 (blue). The adjusted cluster model, shown in Fig. 8, is formed by 12 clusters and has an improved silhouette coefficient of 0.597. This modified clustering more appropriately represents a user's "vision" of the scenario, and can yield better classification results, as discussed next.

6.2. Identifying the relationships between S4 indices and other variables (Q2) and relevant variables to scintillation (Q3)

Proceeding from the previous study and in order to investigate how the multiple tracked variables relate with the S4 indices, we compared the results of classifiers on the set of observations described by three alternative subspaces of variables, taking $S4_corrected$ as target variable and the Tiwari categorization as ground truth. The three sub-spaces considered for describing the observations are formed by (a) all 58 monitored variables that are not directly derived or theoretically associated with measures

Table 3

Classification accuracy of observations described by three distinct variable subspaces.

Technique	58 var. (a)	8 var. (b)	12 var. (c)
J48	96.43%	97.50%	97.69%
Mult. Percep.	96.42%	97.52%	97.85%

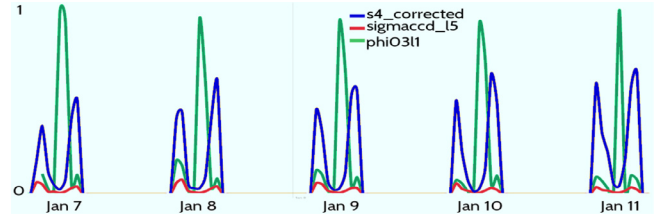


Fig. 9. A five-day sample (from January 7 to 11, 2014, 00:00 h–11:00 h) of the normalized time series of aggregated values (maximum over 1 h intervals) for two representative variables $sigmaccd_{l5_e5b}$ and $phi0311$, and the variable target $s4_corrected$, where a daily pattern is clearly observed.

of the S4 index (obtained discarding the 12 variables included in Cluster C4, depicted in Fig. 7); (b) the representative variables defined by the cluster medoids resulting from X-means (eight variables) clustering, also depicted in Fig. 7; and (c) the representative variables (medoids) defined by the user-adjusted cluster model, shown in Fig. 8, obtained after observing and interacting with the visualizations.

The X-means clustering of all 70 variables (Fig. 7) generated eight clusters, and their corresponding medoids are the variables: si_{l5_e5b} , $dtec_{6045}$, $phi0311$, wn , $azim$, tow , $avgccd_{l2c_e5a}$ and $numerator_{si_{l2c_e5a}}$. The user-adjusted clustering (Fig. 8) has twelve clusters, with medoids: si_{l5_e5b} , $dtec_{6045}$, $phi0311$, wn , $azim$, $avgccd_{l2c_e5a}$, $numerator_{si_{l2c_e5a}}$, $sigmaccd_{l5_e5b}$, $f2nd_tec_locktime$, avg_cn0_{l2} , p_{l5_e5b} and t_{l2c_e5a} .

Classifier accuracy is summarized in Table 3, confirming that both subspaces of variables derived with X-Means and user-adjusted are indeed appropriate to characterize scintillation behavior. According to the scintillation experts, this is a relevant result, as e.g., the eight variables from subset (b) could be monitored using conventional geodesic receivers available, for example, from the Brazilian Network for Continuous Monitoring of the GNSS Systems (RBMC), rather than the highly specialized and more costly ISMRs. In other words, it would be possible to track scintillation with good precision monitoring a reduced number of variables that may be tracked with cheaper receivers.

The behavior of such variables in relation to the S4 indices can be further investigated by studying the normalized line graphs of the aggregated data represented in the time matrices. For instance, Fig. 9 shows a sample of the normalized time series describing two representative variables $sigmaccd_{l5_e5b}$ and $phi0311$, and the target variable $s4_corrected$ during five days in January 2014, where daily patterns are clearly observable. Patterns are also noticeable when observing graphs depicting other representative variables, considering other choices of aggregation function, time spans and time intervals (not shown).

6.3. Comparing the behavior of observations from multiple satellites (Q4)

Identifying groups of satellites more susceptible to scintillation over a region is particularly relevant to derive specific strategies for dealing with this effect on GNSS positioning. A typical task would be concerned with inspecting scintillation values recorded from multiple GPS satellites, and observe their consistency. This

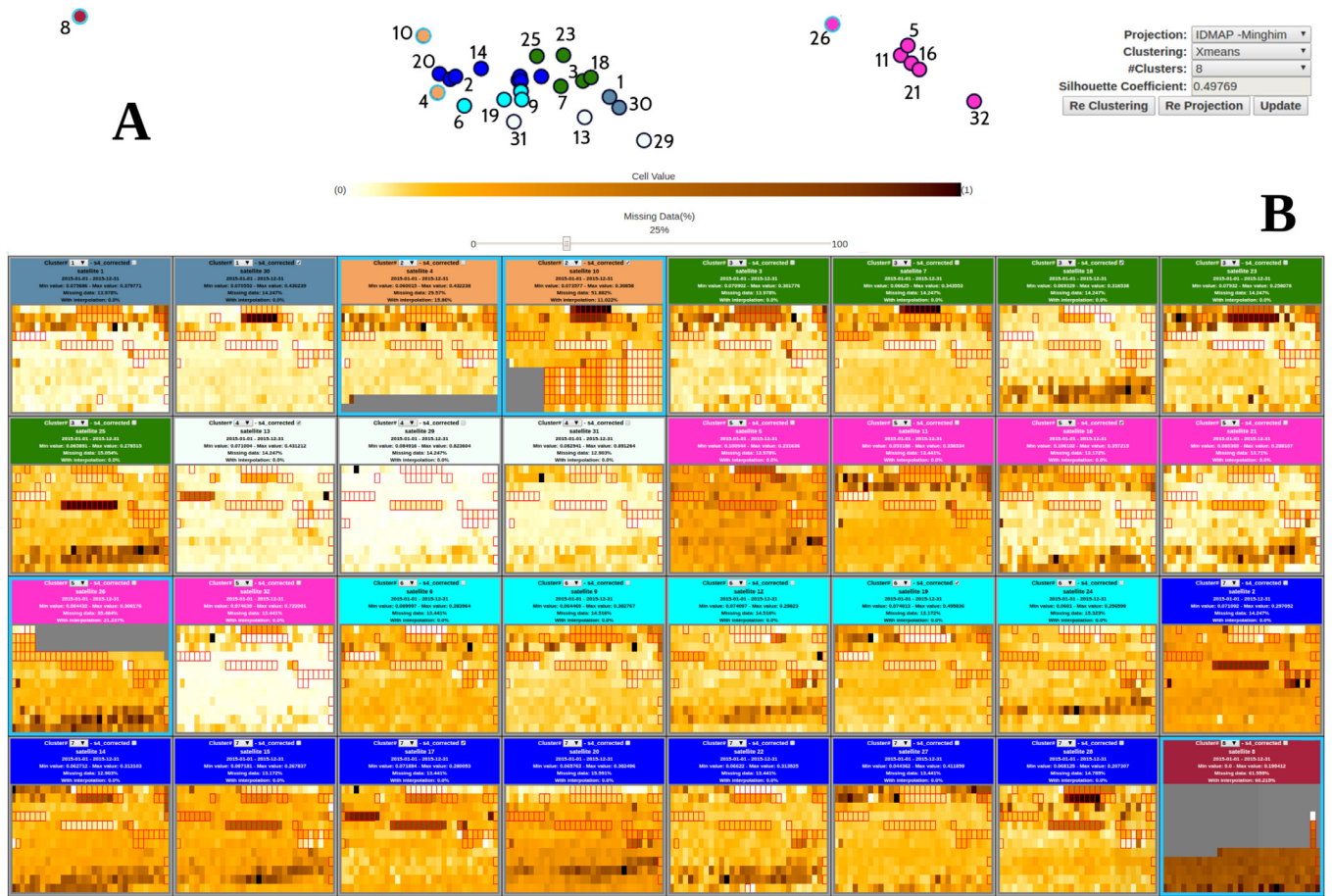


Fig. 10. (A) IDMAP similarity map and (B) Small Multiple Time Matrices view of values of index *S4_corrected* recorded at the 32 GPS satellites (January 1 - December 31, 2015). Matrix cells show average values over daily periods along each month. Numeric labels in the similarity map identify the satellites.

analysis may consider, for example, the *S4_corrected* index, which measures amplitude scintillation computed without ambient noise effects.

We considered observations of the *S4_corrected* index recorded at the PRU2 station from January to December 2015, which were aggregated by taking the average value recorded over daily intervals. This query resulted in 1,775,039 observations from the 32 GPS satellites, which were K-Means clustered picking $k = 8$ (which yielded the best silhouette coefficient, equal to 0.498, from all cluster models obtained with k in [2,15]). The resulting clustering (C1 to C8 identified by colors) is shown in the small multiples time matrix view in Fig. 10.

These clusters reveal distinct patterns of amplitude scintillation, which is typically more severe from November to March. This is confirmed by the time matrices, where the initial and/or final rows (corresponding to the first and last months of the year) show higher averages (darker colors). The three satellites in the white cluster were seemingly less affected by scintillation over this period. The affected satellites at a particular GPS station will change according to the spatial dynamics of both their fixed orbital periods and the ionospheric irregularities. Time matrix views of four satellites (4, 8, 10, 26) depicting over 28% missing data are highlighted (cyan borders). Occurrence of missing data may be related to scintillation susceptibility of a certain satellite because strong scintillation may cause loss of lock and, as a consequence, inability to track some variables. If this is the case, data is missed over short periods (typically a few minutes). One notices that these four satellites display extensive periods of missing data, more likely a

consequence of satellite unavailability, e.g., due to maintenance or replacement operations in the GPS constellation.

7. Conclusions and future work

We have introduced and described an interactive visualization system for exploratory data analysis of ionospheric scintillation data, which are multidimensional and time-varying. Our integrative system approach supports the characterization of temporal relationships between variables associated with the scintillation phenomenon. Scintillation scientists can use our system to pose queries for a target time period and assisted by multiple visualizations integrated with analytical algorithms for clustering and classification can interact with the data and visualizations to find answers to their questions and formulate new scientific hypotheses. Our system is an additional and complementary system to currently available systems [36].

We have presented detailed case studies, conducted in partnership with our collaborating domain experts (co-authors of this paper) to provide supporting scenarios where the system can be employed effectively to discover relevant information about the scintillation phenomenon. Our integrated visual analytics approach allows a user to identify different clusters of variables with quite similar temporal behavior over a target period. We have shown that the subspace defined by the set of medoids, being representatives of each group, suffice to characterize the behavior of the phenomenon. These representative variables show periodic patterns that can be assessed at different time scales. We also explored the

variable *S4_corrected* tracked by the 32 GPS satellites, identifying groups of satellites in which this particular variable exhibits similar behavior and also outliers that could be of interest to the domain specialists. The case studies are illustrative of how the system can assist experts in gaining further insight into the historical behavior of scintillation. Additional experiments and studies of observations collected at other stations over longer time periods will be conducted to validate and further refine the described framework.

We plan to provide similarity maps of observations linked with variable visualizations, making it possible to perform interactions to identify and interpret relevant patterns. As a simple approach linear interpolation can be used, but it would be interesting to investigate possibly other, higher-order interpolation schemes and analyze how they impact the data analysis [49].

We will address scalability of our approaches and their implementation to ensure real-time interactive performance for very large data sets and complex queries. Finally, the proposed analytics strategies are applicable to other domains where large time-varying multivariate data arise and must be understood. Specifically, any multivariate observational data set recorded periodically can be explored with the described visualizations when there is a need to understand and correlate variable subspaces and temporal variation. We are currently generalizing the system and will apply it to stock market and air quality data sets.

Acknowledgments

This work received financial support of the State of São Paulo Research Foundation (FAPESP) grants 11/22749-8 and 17/05838. Aurea Soriano-Vargas acknowledges the support of FAPESP grants 12/24537-0 and 15/12831-0. Bruno C. Vani acknowledges the support of the Federal Institute of Education, Science and Technology of São Paulo (IFSP) and the Coordination for the Improvement of Higher Education Personnel (CAPES) grant 88881.134266/2016-01. Maria Cristina F. Oliveira acknowledges the support of the Brazilian Federal Research Council (CNPq) grant 305696/2013-0.

References

- [1] Yeh K, Liu C. Radio wave scintillations in the ionosphere. *Proc IEEE* 1982;70(4):324–60.
- [2] Alfonsi L, Wernik AW, Materassi M, Spogli L, Bougard B, Monico JFG. Low latitude scintillations: a comparison of modeling and observations within the cigala project. In: Proceedings of the 30th URSI general assembly and scientific symposium; 2011. p. 1–4.
- [3] Kieft P, Aquino M, Dodson A. Using ordinary kriging for the creation of scintillation maps. In: Mitigation of ionospheric threats to GNSS: an appraisal of the scientific and technological outputs of the TRANSMIT project. *InTech*; 2014.
- [4] Vani BC, Shimabukuro MH, Monico JFG. Visual exploration and analysis of ionospheric scintillation monitoring data: the ISMR query tool. *Comput Geosci* 2017;104:125–34.
- [5] Sreeja V, Aquino M, Forte B, Elmas Z, Hancock C, De Franceschi G, et al. Tackling ionospheric scintillation threat to GNSS in latin america. *J Space Weather Space Clim* 2011;1(1):A05.
- [6] Jiao Y, Morton Y. Comparison of the effect of high-latitude and equatorial ionospheric scintillation on GPS signals during the maximum of solar cycle 24. *Radio Sci* 2015;50(9).
- [7] Rezende L, de Paula E, Stephany S, Kantor I, Muella M, de Siqueira P, et al. Survey and prediction of the ionospheric scintillation using data mining techniques. *Space Weather* 2010;8(6):1–10.
- [8] Jäckle D, Fischer F, Schreck T, Keim DA. Temporal mds plots for analysis of multivariate data. *IEEE Trans Vis Comput Graph* 2016;22(1):141–50.
- [9] Kalousis A, Prados J, Hilario M. Stability of feature selection algorithms: a study on high-dimensional spaces. *Knowl Inf Syst* 2007;12(1):95–116.
- [10] Alpaydin E. Introduction to machine learning. MIT Press; 2014.
- [11] Razente HL, Chino FJT, Barioni MCN, Traina AJ, Traina Jr C. Visual analysis of feature selection for data mining processes. In: Proceedings of Brazilian symposium databases (SBBD); 2004. p. 33–47.
- [12] Seo J, Shneiderman B. A rank-by-feature framework for unsupervised multidimensional data exploration using low dimensional projections. In: Proceedings of the IEEE symposium on information visualization; 2004. p. 65–72.
- [13] Piringer H, Berger W, Hauser H. Quantifying and comparing features in high-dimensional datasets. In: Proceedings of the IEEE 12th international conference on information visualization; 2008. p. 240–5.
- [14] May T, Bannach A, Davey J, Ruppert T, Kohlhammer J. Guiding feature subset selection with an interactive visualization. In: Proceedings of the IEEE conference on visual analytics science and technology; 2011. p. 111–20.
- [15] Martínez MJ, Ponzoni I, Díaz MF, Vazquez GE, Soto AJ. Visual analytics in cheminformatics: user-supervised descriptor selection for QSAR methods. *J Cheminform* 2015;7(1).
- [16] Schreck T, Fellner D, Keim D. Towards automatic feature vector optimization for multimedia applications. In: Proceedings of ACM symposium on applied computing; 2008. p. 1197–201.
- [17] Turkay C, Lundervold A, Lundervold AJ, Hauser H. Representative factor generation for the interactive visual analysis of high-dimensional data. *IEEE Trans Vis Comput Graph* 2012;18(12):2621–30.
- [18] Yuan X, Ren D, Wang Z, Guo C. Dimension projection matrix/tree: interactive subspace visual exploration and analysis of high dimensional data. *IEEE Trans Vis Comput Graph* 2013;19(12):2625–33.
- [19] Liu S, Maljovec D, Wang B, Bremer P-T, Pascucci V. Visualizing high-dimensional data: advances in the past decade. *IEEE Trans Vis Comput Graph* 2015a;23(3):1249–68.
- [20] Livingston MA, Decker JW, Ai Z. Evaluating multivariate visualizations on time-varying data. In: Proceedings of the conference on visualization and data analysis, vol. 8654; 2013 p. 86540N.
- [21] Inselberg A, Dimsdale B. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In: Proceedings of the 1st conference on visualization; 1990. p. 361–78.
- [22] Lee T-Y, Shen H-W. Visualization and exploration of temporal trend relationships in multivariate time-varying data. *IEEE Trans Vis Comput Graph* 2009;15(6):1359–66.
- [23] Akibay H, Ma K-L. A tri-space visualization interface for analyzing time-varying multivariate volume data. In: Proceedings of the 9th Joint Eurographics/IEEE VGTC conference on visualization; 2007. p. 115–22.
- [24] Tufte ER. Envisioning information. *Opto Vis Sci* 1991;68(4):322–4.
- [25] Liu X, Hu Y, North S, Shen H-W. Correlated multiples: spatially coherent small multiples with constrained multi-dimensional scaling. *Comput Graph Forum* 2015(1467-8659). doi:10.1111/cgf.12526.
- [26] Zhang D, Zhu L, Wang C, Zhang L. Timespiral, an enhanced interactive visual system for time series data. In: Proceedings of the 2nd international conference on information management (ICIM); 2016. p. 127–33.
- [27] Dasgupta A, Kosara R, Gosink L. Vimtex: a visualization interface for multivariate, time-varying, geological data exploration. *Comput Graph Forum* 2015;vol. 34:341–50.
- [28] Steed CA, Halsey W, Dehoff R, Yoder SL, Paquit V, Powers S. Falcon: visual analysis of large, irregularly sampled, and multivariate time series data in additive manufacturing. *Comput Graph* 2017;63:50–64.
- [29] Li J, Zhang K, Meng Z-P. Vismate: interactive visual analysis of station-based observation data on climate changes. In: Proceedings of the IEEE conference on visual analytics science and technology (VAST); 2014. p. 133–42.
- [30] Poco J, Dasgupta A, Wei Y, Hargrove W, Schwalm C, Cook R, et al. Similarity-explorer: a visual inter-comparison tool for multifaceted climate data. *Comput Graph Forum* 2014;33:341–50.
- [31] Van Dierendonck A, Klobuchar J, Hua Q. Ionospheric scintillation monitoring using commercial single frequency c/a code receivers. In: Proceedings of the ION GPS, vol.93; 1993. p. 1333–42.
- [32] Tiwari R, Skone S, Tiwari S, Strangeways H. Wbmod assisted PLL GPS software receiver for mitigating scintillation affect in high latitude region. In: Proceedings of the 30th URSI general assembly and scientific symposium; 2011. p. 1–4.
- [33] Lima G, Stephany S, Paula E, Batista I, Abdu M, Rezende L, et al. Correlation analysis between the occurrence of ionospheric scintillation at the magnetic equator and at the southern peak of the equatorial ionization anomaly. *Space Weather* 2014;12(6):406–16.
- [34] Lima G, Stephany S, Paula E, Batista I, Abdu M. Prediction of the level of ionospheric scintillation at equatorial latitudes in Brazil using a neural network. *Space Weather* 2015;13(8):446–57.
- [35] Ackah J-B, Obrou O, Groves K. Study of the ionospheric scintillation and tec characteristics at solar minimum in a west african equatorial region using global positioning system (gps) data. In: Proceedings of the 30th URSI general assembly and scientific symposium; 2011. p. 1–4.
- [36] Vani B, Galera Monico J, Shimabukuro M, Pereira V, Aquino M. Exploring the cigala/calibra network data base for ionosphere monitoring over Brazil. In: Proceedings of AGU fall meeting abstracts, vol.1; 2013.
- [37] Warnant R, Pottiaux E. The increase of the ionospheric activity as measured by GPS. *Earth Planet Space* 2000;52(11).
- [38] Liu H, Shu B, Xu L, Qian C, Zhang R, Zhang M. Accounting for inter-system bias in DGNS positioning with GPS/glonass/BDS/galileo. *J Navig* 2017;70(4):686–98.
- [39] Levkowitz H. Color theory and modeling for computer graphics, visualization, and multimedia applications, 402. Springer Science & Business Media; 1997.
- [40] Samara T. Design elements: a graphic style manual. Rockport Publishers; 2007.
- [41] Pelleg D, Moore AW. X-means: extending k-means with efficient estimation of the number of clusters. In: Proceedings of the 17th international conference on machine learning; 2000. p. 727–34.
- [42] Rousseeuw P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1987;20(1):53–65.
- [43] Tejada E, Minghim R, Nonato LG. On improved projection techniques to support visual exploration of multi-dimensional data sets. *Inf Vis* 2003;2(4):218–31.

- [44] Minghim R, Paulovich F, de Andrade Lopes A. Content-based text mapping using multi-dimensional projections for exploration of document collections. In: Proceedings of the international society for optics and photonics (SPIE), visualization and data analysis, 6060; 2006 p. 60600S.
- [45] Faloutsos C, Lin K-I. FastMap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. ACM SIGMOD international conference on Management of data 1995;24(2):163–74.
- [46] Paulovich F, Minghim R. Hipp: a novel hierarchical point placement strategy and its application to the exploration of document collections. IEEE Trans Vis Comput Graph 2008;14(6):1229–36.
- [47] Paulovich F, Nonato L, Minghim R, Levkowitz H. Least square projection: a fast high-precision multidimensional projection technique and its application to document mapping. IEEE Trans Vis Comput Graph 2008;14(3):564–75.
- [48] Jolliffe I. Principal component analysis. Wiley Online Library; 2002.
- [49] Boyles S. Comparison of interpolation methods for missing traffic volume data. In: Proceedings of the transportation research board 90th annual meeting; 2011. 11–3757