Visual Exploration of Ionosphere Disturbances for Earthquake Research

Fan Hong
Peking University

Siming Chen
Peking University

Hanqi Guo
Argonne National Laboratory

Xiaoru Yuan
Peking University

Jian Huang
University of Tennessee, Knoxville

Yongxian Zhang*
China Earthquake Networks Center

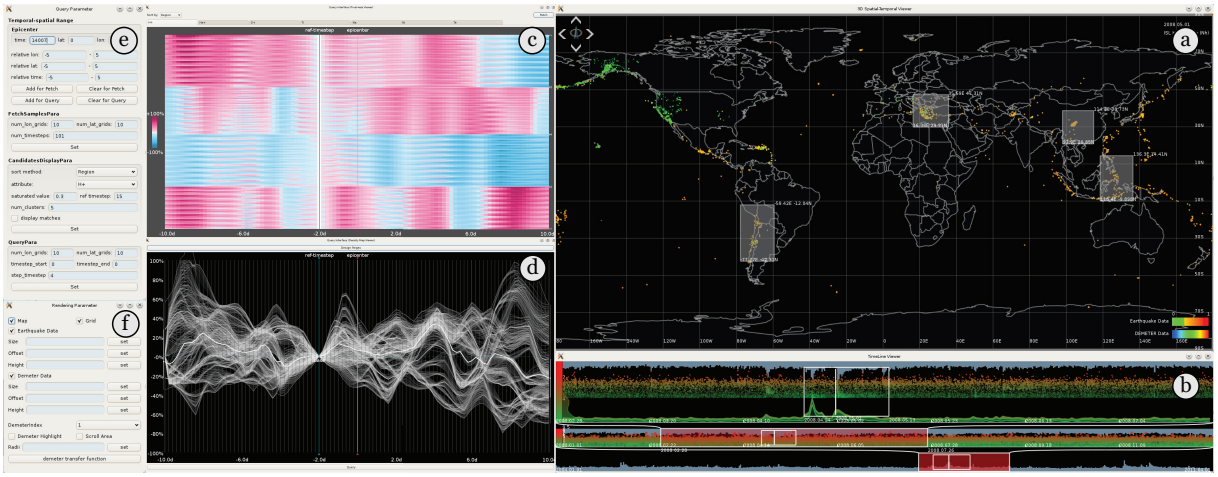

Figure 1: Interface of our system. (a) Map view and (b) timeline view provide a unified exploration space for the heterogeneous dataset. (c) Matrix view and (d) curve view are used to extract disturbance patterns from example DEMETER time-series. (e)(f) Parameter control panels for query, including spatial and temporal extent and resolution, DEMETER attributes, etc.

ABSTRACT

In seismic research, a hypothesis is that ionosphere disturbances are related to lithosphere activities such as earthquakes. Domain scientists are urgent to discover disturbance patterns of electromagnetic attributes in ionosphere around earthquakes, and to propose related hypotheses. However, the workflow of seismic researchers usually only supports pattern extraction from a few earthquakes. To explore the pattern-based hypotheses on a large spatiotemporal scale meets challenges, due to the limitation of their analysis tools. To tackle the problem, we develop a visual analytics system which not only supports pattern extraction of the original workflow in a way of dynamic query, but also extends the work with hypotheses exploration on a global scale. Domain scientists can easily utilize our system to explore the heterogeneous dataset, and to extract patterns and explore related hypotheses visually and interactively. We conduct several case studies to demonstrate the usage and effectiveness of our system in the research of relationships between ionosphere disturbances and earthquakes.

CCS CONCEPTS

• **Human-centered computing** → **Visual analytics**; **Visualization systems and tools**;

KEYWORDS

visual analytics, earthquake, query-based

ACM Reference Format:
Fan Hong, Siming Chen, Hanqi Guo, Xiaoru Yuan, Jian Huang, and Yongxian Zhang. 2017. Visual Exploration of Ionosphere Disturbances for Earthquake Research. In *Proceedings of SA '17 Symposium on Visualization.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3139295.3139301

1 INTRODUCTION

Long-term effort has been taken in seismic community to understand the process of earthquake generation and subsequently to reduce the damage and casualty caused by strong shocks. Recently, observation and research have shown that anomalous electromagnetic signals may have association with seismic activities. One hypothesis is that before earthquakes occur, the anomalies of structures in the epicenter and its tectonic zone can lead to electromagnetic radiance releasing into the air, together with other acoustic and

*e-mails: fan.hong@pku.edu.cn, csm@pku.edu.cn, hguo@anl.gov, xiaoru.yuan@pku.edu.cn, huangj@eecs.utk.edu, yxzhseis@sina.com

SA '17 Symposium on Visualization, November 27-30, 2017, Bangkok, Thailand

ACM ISBN 978-1-4503-5411-0/17/11.
https://doi.org/10.1145/3139295.3139301

geochemical effects. Such effects can introduce seismo-ionosphere phenomena of disturbances in the ionosphere. DEMETER is the first satellite devoted to the study of the ionosphere disturbances related to seismic and human activity, as well as the pre- and post-seismic effects in the ionosphere.

The study on ionosphere disturbances usually involves two kinds of heterogeneous data, i.e. the earthquake catalog data and the DEMETER data. Earthquake catalog data record information about seismic events, e.g. location, magnitude, etc. DEMETER data collect several electromagnetic attributes in the atmosphere with timestamps and locations. Analysis on the heterogeneous dataset can help domain scientists understand the spatiotemporal relationship between ionosphere disturbances and earthquakes.

Currently, dozens of seismic works have been published with their discoveries [Akhoondzadeh et al. 2010; Zhang et al. 2012]. Most of them are on a case-by-case investigation with geographical and statistic tools. Nevertheless, the workflow of case-by-case investigations is limited in some aspects due to their analysis tools. The most important limitation is that the workflow does not consider the popularization of proposed hypotheses to more other earthquakes, which will lower their value. Besides, because domain scientists have no efficient tools to explore the parameter space of patterns, the disturbance patterns only come from direct observation and statistics, which may be hard to reveal the intrinsic relationships. Moreover, these patterns are usually described coarsely and imprecisely, due to the lack of a formal formulation.

In this work, we propose a spatiotemporal visual analytics system, using query-driven approach to extract disturbance patterns, to propose pattern-based hypotheses, and to explore hypotheses on a global scale. Integrated with the original workflow, our system provides a search-by-example way to assist domain scientists to extract disturbance patterns from earthquakes. Moreover, our system has the capability to explore hypotheses in a long time range and on a global scale. Our system can help scientists to improve their research efficiency and gain deeper insights on their research. Our contributions can be summarized as follows:

- We develop a visual analytics system to assist domain scientists to study relationships between earthquakes and ionosphere disturbances.
- We propose a workflow to support hypotheses generation, exploration, and visualization in general seismic research.
- Query-driven approach is utilized to explore the relationships among heterogeneous spatiotemporal datasets.

2 RELATED WORK

In this section, we first review literature on spatiotemporal visualization and query-driven visualization on time-series. Then we discuss the seismic-related visualization works.

Spatiotemporal Visualization. Both earthquake and observational data are spatiotemporal data with multivariate information. Lots of visualization are proposed to visualize spatiotemporal data. Space-time cube [Hägerstraand 1970] is a basic, yet intuitive 3D visualization metaphor. To encode multivariate data, informative glyphs are used in the spatiotemporal scene [Tominski et al. 2012]. To avoid visual clutter brought by 3D visualization, some works directly encode time information [Andrienko and Andrienko 2011; Shanbhag et al. 2005] as well as multivariate information [Bak et al. 2009] on a 2D map.

To further facilitate exploration and analysis, filtering [Krüger et al. 2013], clustering [Andrienko et al. 2009] and aggregation [Andrienko and Andrienko 2008] methods are applied to spatiotemporal data. VIS-STAMP [Guo et al. 2006] involves techniques such as linked parallel coordinates, SOM to analyze multivariate information of spatiotemporal data. Besides, a toolkit proposed by Maciejewski et al. [Maciejewski et al. 2011] supports prediction on hotspots regions with statistical techniques. More recently, discussion on heterogeneous spatiotemporal data analysis is emerging. VASA system [Ko et al. 2014] proposed a comprehensive pipeline to investigate the correlation among weather, traffic and critical infrastructure simulation.

The goal of our work is to help domain scientists discover relationships between earthquakes and ionosphere disturbances. We provide a unified spatiotemporal scene for the exploration of the heterogeneous dataset.

Query-driven Visualization on Time-series. Query-driven visualization is effective when dealing with large datasets [Stockinger et al. 2005]. Dynamic query provides direct interaction and instant feedback when the user manipulates time-series [Hochheiser and Shneiderman 2004]. KronoMiner [Zhao et al. 2011] utilized dynamic query in their circular layout to provide flexible visual comparison for time-series. Buono et al. [Buono et al. 2005] proposed pattern search approach on time-series analysis to find similar occurrences, and also used dynamic query to investigate results. There are other works automatically detecting special subsequences using query techniques, for example to find the most unusual time series subsequences based on periodicity [Keogh et al. 2005], or to detect periodicity and peak/pit [Andrienko et al. 2010]. Besides, Hao et al. [Hao et al. 2011] conduct prediction on time-series with peak preserved. In scientific visualization, a globbing-based query language is proposed to search user specific pattern in a large number of time-series [Glatter et al. 2008]. In our system, dynamic query technique is used to help extract patterns from DEMETER time-series with instant feedback from examples. We also propose a regular expression-inspired query language to facilitate pattern search in our heterogeneous spatiotemporal data.

Seismic Visualization. Research on seismic visualization receives more attention in recent years. Researchers have studied visualization on simulation data and field-measured data using 2D maps [Hirahara et al. 2004] and 3D volumes [Heßel et al. 2015; Hsieh et al. 2010]. Besides, catalog data provides detailed information about earthquakes, e.g. location of epicenter, strike time, etc. Yuen et al.[Yuen et al. 2005] proposed automatic clustering and visualization on catalog data, attempting to reveal spatial features and patterns of earthquake events. Dzwinel et. al. [Dzwinel et al. 2005] also applied cluster analysis on the observed and synthetic catalog data to find multi-resolution structures of earthquake patterns. Integrating the catalog data with satellite-based observational data, Yuan et al. [Yuan et al. 2010] proposed to use parallel coordinates and dimension embedding techniques to study the multi-modal data. Visualization of seismic-center distribution can also be done with time-series analysis based on a 3D correlation graph [Naka et al. 2006]. To our knowledge, there is no previous work analyzing DEMETER data and earthquake catalog data with advanced visual

analytics. Domain scientists show urgent needs to analyze their intrinsic relationships in a more efficient way.

3 DATA DESCRIPTION AND PREPROCESSING

In this work, we develop a visual analytics system to analyze and visualize the relationships between earthquake catalog data and DEMETER data. We introduce these two kinds of data at first, then describe briefly about Kriging interpolation we apply to DEMETER data.

3.1 Data Description

DEMETER data were collected by the DEMETER satellite, which was launched on June 29, 2004, and worked on a quasi sun-synchronous circular orbit with an inclination of 98.23° and an altitude of 710 km. Several scientific sensors are conveyed to measure certain electromagnetic signals, e.g. electron density, ion temperature, etc. We illustrate several consecutive tracks within one day in Figure 2(a), where color encodes the value of electron density. Our DEMETER dataset is provided by China Earthquake Networks Center. DEMETER samples in this dataset have twelve attributes, including timestamp, longitude, latitude, track number, track direction and seven other electromagnetic attributes. These attributes include density of three kinds of ion, i.e. H^+, He^+ and O^+, and overall ion density N_i, electron density N_e, and temperature of ion and electron, T_i and T_e, respectively. The satellite samples these attributes about every three seconds. Figure 2(c) shows the electron density along one track. The satellite worked from Oct. 2004 to Dec. 2010 with a total data size of 15GB.

The earthquake catalog dataset is obtained from the Advanced National Seismic System catalog, which is hosted by the Northern California Earthquake Data Center. It records the location (latitude and longitude), timestamp, magnitude and depth of focus of earthquakes. In our system, to match DEMETER data, we only consider earthquakes occurred in the same temporal range, which results in 645,619 earthquakes. DEMETER samples and earthquake records are registered in a common geographic coordinate system, so that unified exploration on the heterogeneous dataset is possible.

3.2 Kriging Interpolation for DEMETER Data

DEMETER data was obtained during the satellite circling around the Earth, which made the data highly anisotropic. Figure 2(b) employs a spatiotemporal cube for a closer look at the anisotropy. DEMETER samples are dense along the longitude direction, but very sparse along the latitude direction and time dimension. The sparsity poses challenges in existing studies. Existing literatures have to brush a local region to filter DEMETER samples for temporal analysis. However, since these filtered samples spread over a local region, spatial variation could affect the disturbances analysis. At the same time, the sparsity along the time dimension makes it difficult to conduct either minutely or hourly and daily analysis. Therefore, to tackle the sparsity problem, we utilize interpolation to obtain DEMETER samples for any location and any time.

Our interpolation is based on Kriging interpolation [Krige 1951] which complies with the first law of geography, that is *Everything is related to everything else, but near things are more related than distant things*. The basic idea of Kriging interpolation is to model the relationships between distances and difference values from existing samples. Then, by a linear combination of these samples, the estimated value for any position can be obtained. For more details, please check the original work.

To be specific in our scenario, for a fixed position in 3D space (time, longitude and latitude), two pairs of tracks are nearest to it comparing to all other tracks (Figure 2(b)). Every pair of tracks has the same direction. To capture samples on these four tracks, we need to set appropriate spatiotemporal extent around the position. The time span is set to $\pm 1,600$ minutes, since these four tracks cover about 1,400-1,600 minutes (approximately one day). The longitude range is set to $\pm 25°$, because each pair of tracks has an offset of $\sim 25°$ in longitude. As for the latitude range, since the samples are dense along the latitude direction, we limit it to $\pm 5°$ empirically. The above description shows a basic settings for integration. There is a possibility that the ionospheric changes between day-time and night-time could overwhelm the disturbances cased by seismic activity greatly [Němec et al. 2008]. So we allow Kriging interpolation to be apply for day-time and night-time separately. The interpolation result of attribute N_i at a specific timestamp is shown in Figure 2(d). Since the satellite did not pass by the North and South Poles, there is no values in high-altitude regions.

The interpolated data provides at least no less information than the original data, because Kriging interpolation gives exactly the same values for the positions of existing samples. In our case studies, we have shown our system is able to reproduce discoveries found by existing studies with interpolation. Yet, since the analysis approaches are not the same, the analysis results are not always consistent. But if we choose to accept The First Law of Geography and Kriging interpolation based on it, analysis based on interpolated data should be more reliable, because potential regional variations are eliminated for the analysis in our approach.

4 OVERVIEW

Currently, the workflow of domain scientists is as follows. First, a spatiotemporal extent is spotted out around an earthquake. DEMETER samples in the spatiotemporal extent are then filtered out for analysis. Domain scientists will conduct either minutely or secondly analysis based on the original samples, or hourly and daily analysis with derived statistics such as mean or standard deviation. After multiple case-by-case analyses, domain scientists may find some common disturbance patterns, and then propose some hypotheses which relate them with the earthquakes. However, there are three major challenges domain scientists meet:

- Gaps exist between the visualization of the heterogeneous data, which cost more efforts in analyzing their relationships.
- The parameter space of disturbance patterns lacks sufficient exploration. The patterns usually come from direct observations and statistics, which could be hard to reveal the intrinsic relationships.
- In most seismic literature, the disturbance patterns found in the case-by-case workflow are not explored on other cases, which is less convincing.

We proposed a visual analytics system coupled with query-driven techniques to tackle these challenges. After multiple rounds

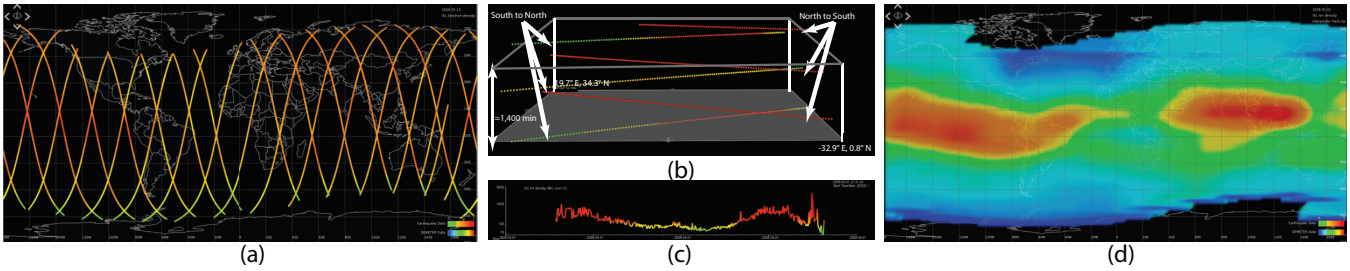

Figure 2: Illustrations of the DEMETER data. (a) Tracks of the DEMETER satellite within one day, color encoding the electron density N_e. (b) A closer illustration of satellite tracks. (c) A line chart to show electron density N_e along one track. (d) Interpolation result of ion density N_i at a specific timestamp.

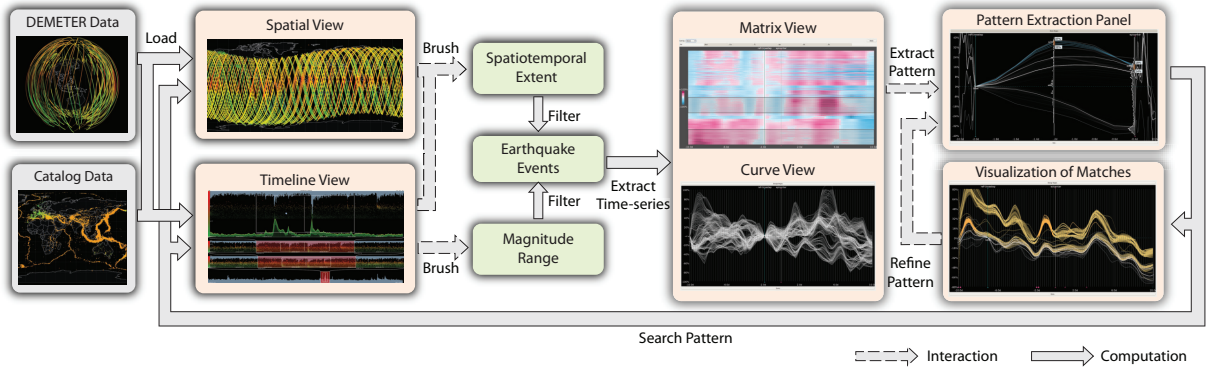

Figure 3: System workflow: spatiotemporal exploration, pattern extraction, hypotheses exploration, and pattern refinement.

of discussion with domain scientists, we summarized the following analysis tasks that our system should provide.

- Support visualization and exploration of the heterogeneous dataset in a unified spatiotemporal scene.
- Integrate the original workflow of pattern extraction into our system in a visual query way, in order to provide more convenient exploration on the parameter space of patterns.
- Extend the original workflow with hypotheses exploration, so that domain scientists can test the hypotheses in a larger set of other earthquakes.

The pipeline of our system is shown in Figure 3. The map view and the timeline view provide a unified scene to visualize and explore the heterogeneous dataset. Earthquake events are selected by interactions in these two views, and then time-series around the epicenters are obtained by Kriging interpolation for pattern extraction. Users extract and formulate patterns with the help of the matrix view and the curve view, and then search patterns in earthquakes on a global scale. The search results are visualized in the map view, timeline view, and the curve view, where users are able to refine patterns and conduct search iteratively.

5 SYSTEM DETAILS

Our system has two major parts: the spatiotemporal exploration part and the pattern extraction and hypotheses exploration part. Working together with these two parts, domain experts are able to generate and explore hypotheses.

5.1 Exploration on the Heterogeneous Dataset

Our system focuses on a heterogeneous dataset, which involves the earthquake catalog data and the DEMETER data. A unified spatiotemporal scene is constructed to connect earthquakes with their possible related DEMETER samples. Users are able to explore their basic relationships in the map view and the timeline view.

Earthquakes and DEMETER samples (within a time span specified by the timeline view) are visualized on a map. We draw points to represent earthquakes (Figure 1(a)) or DEMETER samples (Figure 2(a)). Sizes and colors are used to encode magnitudes of seismic events or ionosphere attributes of DEMETER samples. A continous color scale of green-orange-red is used. In the control panel (Figure 1(f)), users can adjust paramters related with the visualization of map, seismic events, and DEMETER samples. By brushing a region, earthquake events and DEMETER samples are filtered out for the following analysis.

Users can observe the temporal distribution of seimic events, and control the time range for analysis in the timeline view (Figure 1(b)). The x- and y-axis represent the strike time and the magnitude of earthquakes respectively. Earthquakes are drawn with the same color scale as in the map view. Minor earthquakes ($M_s \leq 3$) are aggregated in the theme-river form to avoid clutter. Besides, our timeline view is designed to be multi-level to achieve the scalability in both detail case investigation and long-term pattern analysis.

5.2 Pattern Extraction and Hypothesis Exploration

The most important function of our system is to utilize query-driven techniques to help scientists investigate the relationships between earthquakes and ionosphere disturbances. Our system not only supports the workflow of pattern extraction, but also extends it to explore hypotheses on a global scale (Figure 3). The results of hypotheses exploration are also visualized to provide intuitive feedback about the applicability of hypotheses to help scientists refine disturbance pattern. We will introduce the definition of the pattern at first, then describe these steps in the following subsections.

5.2.1 Definition of Patterns. Domain scientists are interested in certain disturbance patterns which are observed in the evolution of ionosphere attributes. Such disturbance patterns can be literally described as bursting, sharply decrease, consistent increase, and so on. However, this form of description is not accurate for a program to understand. We provide a mathematical definition to describe a *pattern*, which is inspired by the regular expression.

The disturbance *pattern* is defined as a ordered sequence of intervals, each of which represents the lower bound and upper bound of a timestep, e.g. $P \stackrel{\text{def}}{=} [a_1, b_1][a_2, b_2]\cdots[a_m, b_m]$, where $a_i \le b_i$ for every i. For a time-series, $\langle t_0, t_1, \cdots, t_n\rangle$, we say a subsequence $\langle t_p, t_{p+1}, \cdots, t_{p+m}\rangle (0 \le p \le n-m)$ *matches* the pattern P, if the following condition is satisfied: $a_i \le \frac{t_{p+i}}{t_p} - 1 \le b_i$, for $0 < i \le m$. Here we choose relative values instead of absolute values since scientists care more about changes of values instead of original values. $a_{(\cdot)}, b_{(\cdot)}$ can also use a literal value *min* or *max* to represent $\pm\infty$ respectively. To enhance the descriptive capability of the definition, $*$ is allowed to attach to an interval $[a_i, b_i]$, which means the interval can span multiple timesteps. The $*$ makes it convenient to describe disturbance patterns of consistent changes. The definition above acts as an intermediate between users and system, which provides not only readability but also powerful descriptive capability.

5.2.2 Pattern Extraction. Most existing seismic works extract interesting patterns from a single or a few earthquakes, and then propose hypotheses about the relationships between ionosphere disturbances and earthquakes based on discovered patterns. Our system supports this search-by-example workflow of pattern extraction.

Retrieve DEMETER time-series. The extraction starts by selecting earthquakes from catalog data in the map view as described before. A spatiotemporal extent is constructed being centered at each earthquake (both location and time) (Figure 1(a,b)). Then the spatial region is discretized to grids with certain spatial resolution. A time-series is sampled at each grid point (location) using Kriging interpolation with certain temporal resolution. To support different analysis requirements, seismic scientists can adjust parameters such as spatiotemporal extent, spatial resolution, and temporal resolution (Figure 1(e)). These retrieved DEMETER time-series are visualized the matrix view and the curve view for following pattern extraction steps.

Choose time-series of interest in the matrix view. In the matrix view (Figure 4), every row represents one time-series of an ionosphere attribute, while each cell corresponds to one timestep, with color encoding the relative values to a reference timestep. To help analyze group behaviors in time-series, we provide sorting operations according to their geo-locations, their distances to the epicenter, or their mutual similarities. The similarity metric is defined as the accumulation of timestep-wise differences right now. Users can focus on a subset of time-series by brushing to extract disturbance patterns.

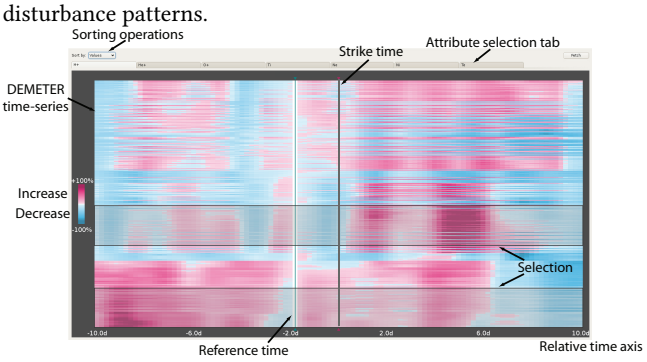

Figure 4: Fetched DEMETER time series are shown in the matrix view, where each row represents one time-series. In the current setting, time-series are ordered based on mutual similarities.

Adjust temporal range of interest in the curve view. The selected time-series are visualized in the curve view (Figure 3 - Curve View) adopting the metaphor of parallel coordinates. Each axis represents one timestep, while position in the axis indicates the relative values to a reference timestep. To avoid severe clutter caused by the large number of time-series, we visualize them in a density-map way, where light gray color indicates high-density area, and vice versa. Based on the behaviors of time-series, scientists can move the reference timestep to where patterns may exist. When moving the reference timestep, we show curves which represent five-number summary (median, 1st/3rd quartile, and minimum/maximum) of every timestep for instant feedback.

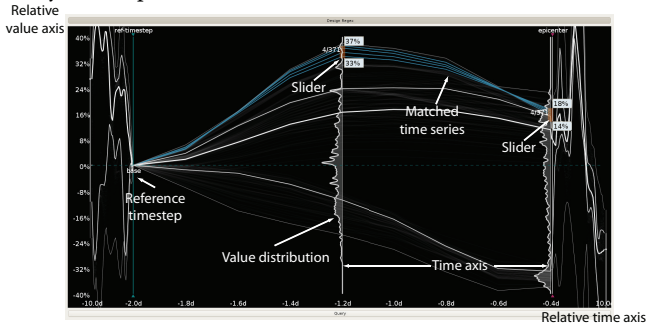

Figure 5: Pattern extraction panel in the curve view. User can extract disturbance patterns by dynamic query with instant feedback from sample time-series.

Formulate patterns by interaction. Users enter the pattern extraction panel (Figure 5) in the curve view. Only a small window of time axes after the reference timestep are focused, while others are squashed. The window represents the *to be extracted* pattern, while each time axis corresponds to one interval in the pattern. Users can freely add an axis to the window or remove one, which indicates an interval is added or removed in the pattern. To adjust the bounds of

each interval, users can drag the slider on the corresponding time axis. Here, we extend the two ends of axes a little to represent *min* and *max* respectively. In addition, users can double-click a time axis to turn on/off the $*$ mark. Using interactions described above, users are expected to extract patterns of interest within the window, and formulate the query using our pattern definition in a visual way.

Our pattern extraction works in a search-by-example way, that is in the whole process, the example time-series always exist as hints to assist users. When dragging the sliders, time-series that match the pattern are highlighted in blue. Besides, the distribution curve of values is attached on each time axis. For each axis, we also show how much example time-series have passed the partial pattern which spans from the first axis till itself. All these features are designed to help scientists better extract the disturbance patterns.

After the disturbance pattern extracted, scientists are able to propose certain hypotheses which relate the properties of earthquake events with the disturbances patterns.

5.2.3 Hypotheses Exploration on a Global Scale. After we extract disturbance patterns, a natural idea is to test the proposed hypotheses on other earthquakes. Current seismic researches are limited due to analysis tools, which are unable to explore hypotheses on a larger spatial and temporal scale. In our system, users are allowed to explore hypotheses on earthquakes occurred all over the world. Firstly, the system filter out earthquakes based on users' setting, e.g. magnitude range, temporal range, etc. DEMETER time-series are generated around every earthquake with the same way as described before. Then, the system searches the disturbance pattern in all time-series and obtains the matched locations and timesteps.

The search results are visualized in the map view and timeline view. Earthquakes are colored in yellow or blue to indicate whether pattern is found or not around the epicenters (Figure 7(a)(b)). When clicking on one earthquake, a bar chart is used to show the ratio of matches in every timestep (Figure 7(c)). Here the ratio of matches is defined as the ratio of matched locations in all interpolated grid points. In this way, scientists can observe when the disturbances exist compared to the strike time. Besides, those locations with patterns found are drawn on the map (Figure 7(c)), which provides possible hints to relate the locations of disturbances with earthquakes. At the same time, those matched time-series are overlaid in the curve view using a different color (Figure 3 - Visualization of Matches) which provides hints to refine the disturbance patterns.

6 CASE STUDIES

We conduct three case studies to demonstrate the usage of our system. At first, we focus on a single earthquake and reproduce the exploration process from a previous seismic work. Then, we describe how to use our visual query interface to extract patterns from example DEMETER samples, and how to explore hypotheses on earthquakes around the world. At last, we conduct global exploration of two hypotheses which come from seismic literature, and give some statistical results.

6.1 Analysis on 2008 Wenchuan Earthquake

Wenchuan M_s 8.0 earthquake occurred at 06:28 (UTC) on May 12, 2008 and was located in Sichuan Province (31.0°N, 103.4°E), China. Disturbances in electron density N_e were found near the epicenter (within ±10° in both latitude and longitude) [He et al. 2011]. To be specific, a rectangular region is spotted centered at the epicenter. DEMETER samples in that region are filtered out. From these samples, daily averages are calculated to derive a time-series, where they only consider the time span from 15 days before the earthquake to 15 days after. Those timesteps whose values are beyond one standard deviation of the full-time average are considered as disturbances. In this way, several peaks and valleys are identified. We try to reproduce the exploration process.

A spatial region is spotted which spans from 22.9°N, 94.4°E to 39.7°N, 111.6°E. Then temporal range is set to 15 days before and after the strike time. Then time-series are sampled from the spatiotemporal extent using Kriging interpolation. We then set the disturbance pattern as $[0.10, max][min, 0]$, so as to identify peaks of N_e around the epicenter. We set the temporal resolution to 8 hours for a finer analysis.

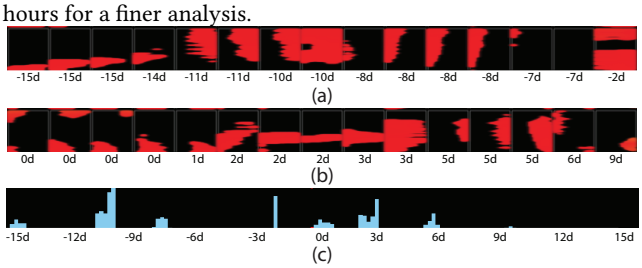

Figure 6: Visualization of pattern search results of attribute N_e for Wenchuan earthquake. Visualization of matched locations at selected timesteps before (a) and after (b) Wenchuan Earthquake happened. (c) A bar chart showing the ratio of matches at each timestep.

Figure 6 shows the visualization results of matched locations and timesteps in both timeline view and map view. In the map view (a)(b), those areas with the disturbances pattern found can be clearly observed. In the timeline view (c), the ratio of matches are visualized as a barchart. From the visualization, we can identify several high peaks of disturbances in 15 days, 11-13 days and 2 days before the earthquake, and in 0-3 days and 5-6 days after the earthquake. Most of these peaks agree with the discoveries in the previous work [He et al. 2011], while some are new in our system which we believe are caused by the differences of the analysis methods.

6.2 Pattern Extraction from Examples and Hypotheses Exploration

Our system is capable of extracting disturbance patterns from example DEMETER time-series, and then exploring these pattern-based hypotheses globally. In this case, we demonstrate the capability.

Users firstly select several earthquakes from the catalog data. The spotted regions are shown in Figure 1(a)), and the temporal range spans from April 24, 2008 to May 14, 2008 (Figure 1(b)). For every earthquake, users set a spatiotemporal extent centered at the epicenter, in this case ±5° for space and ±5 days for time. The system then derives time-series in this extent with interpolation. In the matrix view, these time-series are sorted based on their similarity (Figure 4). One group of time-series is observed to have a peak just after the earthquake happens, and has consistent high values 4 days after the strike time. The other group shows low values in the second days

before the earthquake happens. These two groups of time-series are selected for pattern extraction (Figure 5). In this case, a peak pattern starting from 2 days before the strike is extracted from the latter group, and a regular expression $[0.34, 0.37][0.14, 0.19]$ is formulated to describe such pattern. We search this disturbance pattern in earthquakes which satisfy $5.0 \leq M_s \leq 6.0$ and occurred in the year 2008.

The search results are visualized in both the map view and the timeline view. In the map view (Figure 7(a)), we observe that unmatched earthquakes are mainly distributed in the low-latitude region, especially near the equator. By clicking on earthquakes of interest in the map view or the timeline view, users are able to investigate where and when the disturbance pattern is found, which may help scientists to proposed more specific hypotheses. In Figure 7(c), the earthquake (1) and (3) show very different spatial distribution of the matched locations, and the pattern is found at different timesteps. While for the earthquake (2), it shows strong signals that the disturbance happens just one day before the earthquake happens. Based on the observations, scientists are able relate the properties of earthquake events and the disturbances, such as the time, location, direction, degree, etc. Then, users can refine the query conditions based observations on one example, and reconduct query again to further study if the proposed relationships widely exist for all other earthquake events.

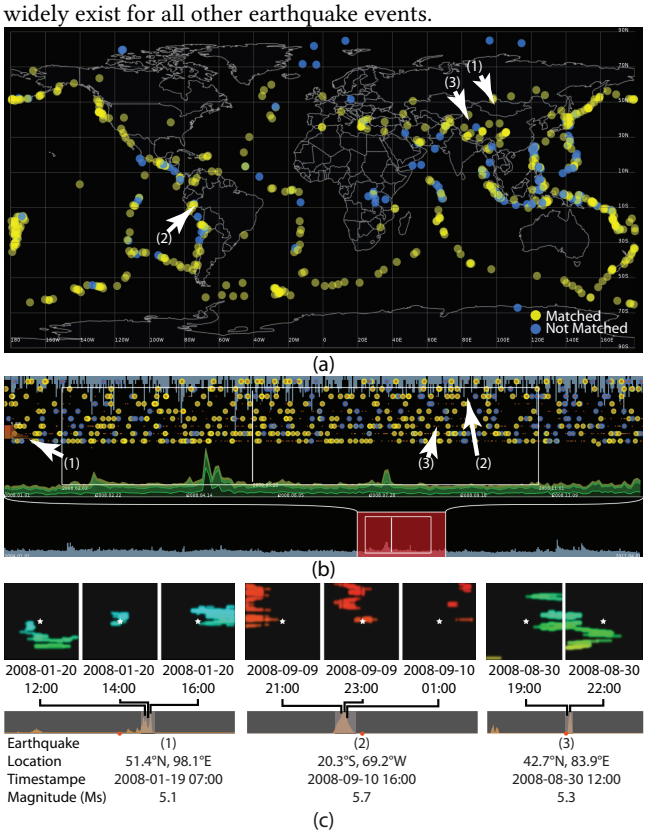

Figure 7: The pattern search results are displayed on the map view (a) and the timeline view (b). 3 events are chosen for a closer investigation of temporal and spatial distributions (c).

6.3 Exploration of Existing Hypotheses

In this case, we explore two existing hypotheses from seismic literature, and to investigate how much these hypotheses related to occurrences of earthquakes. We consider the following two disturbance patterns: peaks in electron density N_e: the disturbance pattern is formulated as $[0.34, 0.37][0.14, 0.19]$, and resolution of time-series is set to 0.8 day; consistent increases in electron density N_e [Zhang et al. 2010], which is found in the 2009 Sumatra earthquake: the pattern is set as $[0.03, 0.07][0.20, 0.23][0.40, max]$. Earthquakes occurred in the year of 2008 are taken into our consideration, and are grouped based on their magnitudes.

Search results for these two patterns are shown in Table 1. We can observe that in more than 60 percent of earthquakes the disturbance pattern of peak is found, while more than 75 percent of earthquakes have the consistent increase pattern. The consistent increase pattern is shown to be more common than the peak pattern. We also found these two disturbance patterns are more common in the group [0-3], especially for the peak pattern. While for the group [6-9.5], these two patterns are less likely to be found.

We should note such statistics do not directly indicate correlation or causality between seismic events and ionospheric disturbances. A full assessment of the hypotheses is still required. Nevertheless, our system provides a powerful tool for domain scientists to gain a deeper understanding of their scientific hypotheses than before.

Table 1: Statistic results of hypotheses exploration.

Pattern 1: peaks					
Mag.	>6	5-6	4-5	3-4	0-3
Matches	68	395	957	714	399
Total	107	579	1413	1078	500
Ratio	63.6%	68.9%	67.6%	66.2%	79.8%
Pattern 2: consistent increases					
Mag.	>6	5-6	4-5	3-4	0-3
Matched	81	465	1129	854	414
Total	107	579	1413	1078	500
Ratio	75.7%	79.9%	80.3%	79.2%	82.8%

7 EVALUATION AND DISCUSSION

We evaluate the system from multiple perspectives. Firstly, the results of Kriging interpolation (Figure 2(d)), as the basis of our system, are confirmed by our domain collaborators. Secondly, we have reproduced the original workflow with a case which has been studied in seismic community before. Most results of our approach agree with those in the previous work. While for our new findings, we think they come from the different analysis methods, especially the Kriging interpolation. Right now, because of the lack of ground truth or auxiliary datasets, we are unable to determine whether they are false-positive or not. But since Kriging interpolation is widely used in geography, we have confidence in our approach. Thirdly, we also received the feedback from our domain collaborators. The unified exploration for the heterogeneous dataset, and the query-based hypotheses exploration were appreciated. They said the visualization of both single earthquake exploration and hypotheses exploration gave very direct spatiotemporal indicators for the relationships between earthquake events and disturbances. Though powerful enough, the pattern extraction function was complained

as complicated to learn. At the same time, the exacted patterns, as well as the following proposed hypotheses based on them, were not fully recognized from statistic results, as we have shown in Figure 7. They actually need further investigations, but our system could not provide adequate guidance. They also suggested to analyze the relationships with multiple attributes simultaneously. We would like to improve these features in the future work, and try to balance the complexity and functionality of our system.

Our system workflow involves lots of parameters to be decided, e.g. the spatial and temporal resolution when extracting time-series around earthquakes (Figure 1(e)). Different parameter settings could influence the analysis results. Currently users can only set these parameters empirically. But in the future, we need to provide formal sensitivity analysis for these parameters to help decide their choices. Yet, for the sensitivity of disturbance patterns, we have provided lots of derived information from example time-series, which can be treated as instant sensitivity feedback.

8 CONCLUSION

In this paper, a visual analytics system is proposed to explore the relationships between ionosphere disturbances and earthquakes. Our system supports pattern extraction from earthquakes using the visual query, which improves the efficiency of the original workflow. Moreover, we extend the workflow with hypotheses exploration on a global scale, where domain scientists can acquire a deeper understanding of the intrinsic relationships between earthquakes and ionosphere disturbances. In the future, we would like to conduct formal sensitivity analysis for parameters, and multivariate analysis. We also want to improve the usability of our system.

ACKNOWLEDGMENTS

The authors wish to thank anonymous reviewers for their comments. This work is supported by NSFC No. 61170204 and is partially supported by NSFC Key Project No. 61232012.

REFERENCES

M. Akhoondzadeh, M. Parrot, and M. R. Saradjian. 2010. Electron and ion density variations before strong earthquakes (/M/>6.0) using DEMETER and GPS data. *Nat. Hazards Earth Syst. Sci.* 10 (2010), 7–18.

Gennady L. Andrienko and Natalia V. Andrienko. 2008. Spatio-temporal aggregation for visual analysis of movements. In *Proc. of IEEE VAST*. 51–58.

Gennady L. Andrienko, Natalia V. Andrienko, Martin Mladenov, Michael Mock, and Christian Pölitz. 2010. Discovering bits of place histories from people's activity traces. In *Proc. of IEEE VAST*. 59–66.

Gennady L. Andrienko, Natalia V. Andrienko, Salvatore Rinzivillo, Mirco Nanni, Dino Pedreschi, and Fosca Giannotti. 2009. Interactive visual clustering of large collections of trajectories. In *Proc. of IEEE VAST*. 3–10.

Natalia V. Andrienko and Gennady L. Andrienko. 2011. Spatial Generalization and Aggregation of Massive Movement Data. *IEEE Trans. Vis. Comput. Graph.* 17, 2 (2011), 205–219.

Peter Bak, Florian Mansmann, Halldor Janetzko, and Daniel A. Keim. 2009. Spatiotemporal Analysis of Sensor Logs using Growth Ring Maps. *IEEE Trans. Vis. Comput. Graph.* 15, 6 (2009), 913–920.

Paolo Buono, Aleks Aris, Catherine Plaisant, Amir Khella, and Ben Shneiderman. 2005. Interactive pattern search in time series. In *Electronic Imaging 2005*. International Society for Optics and Photonics, 175–186.

Witold Dzwinel, David A.Yuen, Krzysztof Boryczko, Yehuda Ben-Zion, Shoichi Yoshioka, and Takeo Ito. 2005. Nonlinear multidimensional scaling and visualization of earthquake clusters over space, time and feature space. *Nonlinear Processes in Geophysics* 12 (2005), 117–128.

Markus Glatter, Jian Huang, Sean Ahern, Jamison Daniel, and Aidong Lu. 2008. Visualizing Temporal Patterns in Large Multivariate Data using Modified Globbing. *IEEE Trans. Vis. Comput. Graph.* 14, 6 (2008), 1467–1474.

Diansheng Guo, Jin Chen, Alan M. MacEachren, and Ke Liao. 2006. A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP). *IEEE Trans. Vis. Comput. Graph.* 12, 6 (2006), 1461–1474.

Torsten Hägerstraand. 1970. What about people in regional science? *Papers in regional science* 24, 1 (1970), 7–24.

Ming C. Hao, Halldór Janetzko, Sebastian Mittelstädt, W. Hill, Umeshwar Dayal, Daniel A. Keim, Manish Marwah, and Ratnesh K. Sharma. 2011. A Visual Analytics Approach for Peak-Preserving Prediction of Large Seasonal Time Series. *Comput. Graph. Forum* 30, 3 (2011), 691–700.

Yufei He, Dongmei Yang, Jiadong Qian, and Michel Parrot. 2011. Anomaly of the ionospheric electron density close to earthquakes: Case studies of PuâĂŹer and Wenchuan earthquakes. *Earthquake Science* 24 (2011), 549–555. Issue 6. 10.1007/s11589-011-0816-0.

Stefan Heßel, Oliver Fernandes, Sebastian Boblest, Philipp Offenhäuser, Malte Hoffmann, Andrea Beck, Thomas Ertl, Colin Glass, Claus-Dieter Munz, and Filip Sadlo. 2015. Visualization of 2D wave propagation by Huygens' principle. In *Proceedings of the 15th Eurographics Symposium on Parallel Graphics and Visualization*. Eurographics Association, 19–28.

K. Hirahara, N. Kato, T. Miyatake, T. Hori, M. Hyodo, J. Inn, N. Mitsui, Y. Wada, T. Miyamura, Y. Nakama, T. Kanai, and M. Iizuka. 2004. *Simulation of Earthquake Generation Process in a Complex System of Faults*. Technical Report. Earth Simulator Center.

Harry Hochheiser and Ben Shneiderman. 2004. Dynamic query tools for time series data sets: Timebox widgets for interactive exploration. *Information Visualization* 3, 1 (2004), 1–18.

Tung-Ju Hsieh, Cheng-Kai Chen, and Kwan-Liu Ma. 2010. Visualizing field-measured seismic data. In *Proc. of IEEE Pacific Visualization Symposium*. 65–72.

Eamonn J. Keogh, Jessica Lin, and Ada Wai-Chee Fu. 2005. HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence. In *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM 2005), 27-30 November 2005, Houston, Texas, USA*. 226–233.

S. Ko, J. Zhao, J. Xia, S. Afzal, X. Wang, G. Abram, N. Elmqvist, L. Kne, D. Van Riper, K. Gaither, S. Kennedy, W. Tolone, W. Ribarsky, and D.S. Ebert. 2014. VASA: Interactive Computational Steering of Large Asynchronous Simulation Pipelines for Societal Infrastructure. *IEEE Trans. Vis. Comput. Graph.* 99 (2014).

D.G Krige. 1951. *A statistical approach to some mine valuations and allied problems at the Witwatersrand*. Master's thesis. University of Witwatersrand.

Robert Krüger, Dennis Thom, Michael Wörner, Harald Bosch, and Thomas Ertl. 2013. TrajectoryLenses - A Set-based Filtering and Exploration Technique for Long-term Trajectory Data. *Comput. Graph. Forum* 32, 3 (2013), 451–460.

Ross Maciejewski, Ryan Hafen, Stephen Rudolph, Stephen G. Larew, Michael A. Mitchell, William S. Cleveland, and David S. Ebert. 2011. Forecasting Hotspots - A Predictive Analytics Approach. *IEEE Trans. Vis. Comput. Graph.* 17, 4 (2011), 440–453.

Takatoshi Naka, Masashi Yamada, Mamoru Endo, Shinya Miyazaki, and Junichi Hasegawa. 2006. Visualization of Seismic-center Distribution Data for Earthquake Prediction. In *Proc. of Nicograph Intl, Session V Simulation*.

František Němec, Ondrej Santolík, Michel Parrot, and Jean-Jacques Berthelier. 2008. Spacecraft observations of electromagnetic perturbations connected with seismic activity. *Geophysical Research Letters* 35, 5 (2008).

Poonam Shanbhag, Penny Rheingans, and Marie desJardins. 2005. Temporal Visualization of Planning Polygons for Efficient Partitioning of Geo-Spatial Data. In *IEEE Symposium on Information Visualization (InfoVis 2005), 23-25 October 2005, Minneapolis, MN, USA*. 28.

Kurt Stockinger, John Shalf, Kesheng Wu, and E. Wes Bethel. 2005. Query-Driven Visualization of Large Data Sets. In *16th IEEE Visualization Conference (VIS 2005), 23-28 October 2005, Minneapolis, MN, USA*. 22.

Christian Tominski, Heidrun Schumann, Gennady L. Andrienko, and Natalia V. Andrienko. 2012. Stacking-Based Visualization of Trajectory Attribute Data. *IEEE Trans. Vis. Comput. Graph.* 18, 12 (2012), 2565–2574.

Xiaoru Yuan, He Xiao, Hanqi Guo, Peihong Guo, Wesley Kendall, Jian Huang, and Yongxian Zhang. 2010. Scalable Multi-variate Analytics of Seismic and Satellite-based Observational Data. *IEEE Trans. Vis. Comput. Graph.* 16, 6 (2010), 1413–1420.

D. A. Yuen, B. J. Kadlec, E. F. Bollig, W. Dzwinel, Z. A. Garbow, and C. R. S. da Silva. 2005. Clustering and Visualization of Earthquake Data in a Grid Environment. *Visual Geosciences* (2005), 1–12.

X. Zhang, X. Shen, M. Parrot, Z. Zeren, X. Ouyang, J. Liu, J. Qian, S Zhao, and Y Miao. 2012. Phenomena of electrostatic perturbations before strong earthquakes (2005-2010) observed on DEMETER. *Natural Hazards and Earth System Science* 12, 1 (2012), 75–83.

Xue-Min Zhang, LIU Jing, SHEN Xu-Hui, M Parrot, QIAN Jia-Dong, Xin-Yan OUYANG, ZHAO Shu-Fan, and Jian-Ping HUANG. 2010. Ionospheric perturbations associated with the M8. 6 Sumatra earthquake on 28 March 2005. *Chinese Journal Geophysics* 53, 3 (2010).

Jian Zhao, Fanny Chevalier, and Ravin Balakrishnan. 2011. KronoMiner: using multi-foci navigation for the visual exploration of time-series data. In *Proc. of ACM CHI*. 1737–1746.