Article

Association Rules-Based Multivariate Analysis and Visualization of Spatiotemporal Climate Data

Feng Wang [1,2], Wenwen Li [1,*], Sizhe Wang [1] and Chris R. Johnson [2]

[1] School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85287-5302, USA; fwang911@gmail.com (F.W.); wsizhe@asu.edu (S.W.)

[2] Scientific Computing and Image Insititute, University of Utah, Salt Lake City, UT 84112, USA; crj@sci.utah.edu

*   Correspondence: wenwen@asu.edu; Tel.: +1-480-727-5987

Abstract: Understanding atmospheric phenomena involves analysis of large-scale spatiotemporal multivariate data. The complexity and heterogeneity of such data pose a significant challenge in discovering and understanding the association between multiple climate variables. To tackle this challenge, we present an interactive heuristic visualization system that supports climate scientists and the public in their exploration and analysis of atmospheric phenomena of interest. Three techniques are introduced: (1) web-based spatiotemporal climate data visualization; (2) multiview and multivariate scientific data analysis; and (3) data mining-enabled visual analytics. The Arctic System Reanalysis (ASR) data are used to demonstrate and validate the effectiveness and usefulness of our method through a case study of "The Great Arctic Cyclone of 2012". The results show that different variables have strong associations near the polar cyclone area. This work also provides techniques for identifying multivariate correlation and for better understanding the driving factors of climate phenomena.

1. Introduction

Over the past few decades, enhanced Earth observation techniques and powerful climate modeling capabilities have facilitated the exponential growth of spatiotemporal climate data [1]. These data carry important information for discovering and analyzing the underlying pattern of various atmospheric phenomena. To gain an in-depth understanding of the dynamic mechanics of the extinction and reignition of these phenomena, such as cyclones, scientists need intuitive and convenient approaches to validate known associations and reveal hidden ones in multiple physical properties [2]. There is an increasing need for methods and systems to support heuristic data exploration based on interactive visualization and analytics. However, designing an effective and easy-to-use heuristic climate visualization system presents a variety of challenges.

First, high dimensions, heterogeneity, time variation, and large volumes characterize spatiotemporal climate data [1]. Effectively simulating and visualizing these multi-dimensional and multivariate data are challenging. Second, atmospheric phenomena, which span time and lack concrete form, usually have fuzzy boundaries. The patterns of these phenomena are often hidden in the structured data and cannot be easily revealed with traditional visual interaction (such as value picking and filtering). Moreover, the users of visualization systems have different backgrounds, skills, interests, and needs [3]. To address these challenges, sophisticated visualization techniques, well-designed user interfaces, and data mining algorithms need to be seamlessly integrated to exploit effective visual exploration and scientific discovery.

Previous studies on developing climate data visualization systems have bridged the gap between climate scientists and the massive amounts of data they study [4–6]. For instance, the Visualization and Analysis Platform for Ocean, Atmosphere, and Solar Researchers (VAPOR) [7] provided by the National Center for Atmospheric Research (NCAR) depicts climate datasets as comprehensible visual representations. ParaView [8] benefits users from a range of scientific fields with a general-purpose, open-source, visual analytic framework. UV-CADT [9] is an open-source, desktop-based tool for climate visualization and analysis. Although providing an intuitive visual exploration interface, these systems are limited in supporting space-time analysis if they are not equipped with the ability to integrate data into virtual geographical environment, such as a digital earth or a virtual globe. Moreover, in existing tools, users have difficulty in selecting the principal variables related to a specific phenomenon and understanding the associations between these and other variables [10]. To this end, a heuristic climate visualization system is presented to benefit climate scientists and the public by reducing visual analytics obstacles.

The framework introduced in this paper integrates intuitive 4D visualization and mining of hidden relationships among multiple climate variables. It is built upon web infrastructure, enabling easy access to the system at any location without software installation or configuration. The system includes scale visualization tools including volume rendering and vector streamline. Two major components, spatial view and parallel coordinate plot (PCP) view, are integrated into the system to improve human–computer interaction and to assist in the multivariate analysis. In addition, an association rule-learning algorithm is implemented to generate strong association rules among variables and offer insights to end-users. To demonstrate the effectiveness and usefulness of our system, we use a strong polar cyclone as a case study.

There are two main contributions of this work. First, we provide a real-time scalable visualization system for spatiotemporal climate data. Second, we combine visualization techniques with data mining algorithms to facilitate interactive visual exploration of scientific data by various user groups.

2. Related Works

In this section, we summarize the previous related work in (1) multivariate spatiotemporal data visualization and (2) visual analytics of multivariate climate data.

Multivariate spatiotemporal data visualization: Spatiotemporal climate data visualization remains an important and active research topic in scientific visualization. Utilizing visualization techniques to gain insight from the underlying information has been the focus of many research efforts [11]. Previous solutions for visual exploration of multivariate data can be categorized into two classes [12]. The first focuses on developing efficient visualization techniques to facilitate interactive data exploration and providing user-friendly widgets to support the comparison of variables side by side. Those studies vary from producing an image representation of the space-time feature with slicing and projection techniques to designing specifications for a transfer function of a time-varying variable and/or developing various exploration interfaces to facilitate data recognition. For example, Woodring and Shen [13] proposed a method to compare the differences in multivariate, time-varying, and comparative data by combining several volumes from different variables. Wang et al. [14] introduced information theory to enhance visualization and guide framework design. Guo et al. [6] integrated multiple components, including PCP and multi-dimensional scaling plots, to design an interesting transfer function interface. Jankun-Kelly and Ma [15] conducted research to determine how the dynamic behavior of spatiotemporal data can be captured using different summary functions. Janicke et al. [16] introduced an approach that transforms multivariate data to a point cloud organized in a 2D plane, allowing intuitive analyses of many variables together with brushing, scatterplots, and linked views. Potter et al. [17] proposed ensemble-Vis to allow scientists to gain key scientific insight into simulation data as well as the uncertainty associated with the data.

The second solution combines all variables into a single view to explore the space distribution of data and their underlying relationships. For instance, Akiba et al. [2] introduced data fusion strategies

to handle multiple values recorded in the same spatial domain. Luo and Dingliana [18] designed a transfer function in the HSB (hue, saturation, and brightness) color space to view multiple variables in a cyclone simulation. Although useful, these works have some limitations, including the difficulty of comparing multiple variables simultaneously, the lack of quantitative evaluation of multiple variables, and the sacrifice of geospatial characteristics of the original data.

Visual analytics of multivariate climate data: Association analysis—a rule-based machine learning method—has been widely used in data-driven analytics to identify latent relationships among multi-variables. A variety of other approaches, including biclustering techniques [19] and cosine similarity measurements [20], have been used to identify the relationships of multiple variables as well. With the development of machine learning techniques, data mining algorithms are being used to discover the underlying knowledge of climate phenomena. For instance, Li et al. [21] revealed how the formation and intensity of tropical cyclones is related to climate factors, such as surface temperature and water vapor, using a decision tree learning method. Other variations of classic decision tree algorithms have also been used. Catani et al. [22] introduced a specific version of random forest algorithms to evaluate the association of landslide susceptibility and geology and geography variables. Mounce et al. [23] applied genetic algorithms along with numerical regression as a hybrid data mining technique to discover the relationship between accumulation of discoloration material and other information in a drinking water system. Except for symmetric relationships, the directional connection between variables has also been taken into account in recent research. Liu and Shen [10] associated different scalar variables by modeling the directional interaction using a social network model. Yang et al. [24,25] conducted work on predicting tropical cyclone strength using association rule analysis.

Although the research listed above has contributed in specific domains (i.e., tropical cyclone, scientific visualization), no studies have examined and validated data in the polar region, which is and important region of interest in climate research. Meanwhile, very few studies have focused on integrating machine learning algorithms into a visualization system and providing a heuristic strategy to support visual analytics. To this end, we incorporate association rule learning into our visualization framework to provide a quantitative analysis for further understanding of the underlying associations of the polar climatic phenomena. The intuitive and interactive visualization components will facilitate and accelerate the knowledge discovery processes.

3. Spatiotemporal Climate Data

Climate data are characterized by complexity and heterogeneity because of their high dimensionality, multivariability, and multiresolution factors, which make visualizing and analyzing the data challenging. As observation techniques develop, climate data have evolved into four categories: in situ, remotely sensed, model output, and paleoclimate [26]. Currently, climate model simulation and reanalysis datasets are the main contributors to climate data. As the simulation-observation hybrid data, reanalysis datasets have the advantages of data from both observation and numerical climate modeling. Numerical climate modeling provides simulation data covering a large scale with time-series, which are not easily or routinely observed. Meanwhile, because it is calibrated by observation data, the re-analysis data is more accurate than using simulation models alone. Generally, these data have the following characteristics [26]:

- Autocorrelation—climate data that are close in space and time are more similar than data that are farther apart.
- Ambiguous boundaries—spatiotemporal phenomena in climate research are abstract "objects" and evolving patterns over the spatiotemporal span.
- Uncertainty, variability, and diversity—these characteristics stem from the biases in sampling and measurement.
- Multivariability-climate data have different attributes, such as temperature, humidity, wind speed, etc.

- High dimensionality—climate data are observed in 3D space and mapped to long time-stamps, which can vary from months to years.

These characteristics pose a significant challenge to climate scientists in both visualization and data analysis. Autocorrelation, for example, limits the accuracy of models that assume independent and identically distributed observations in climate data. Ambiguous boundaries complicated the data mining process for making accurate extraction in patterns belonging to the core of a phenomenon. However, solutions to this challenge are emerging gradually, which combine visual analysis and data mining. Data mining algorithms can be effectively integrated into the analysis process and significantly benefit the data exploration. Therefore, our goal is to develop effective visual analytical tools to support the knowledge mining of complex, big climate data.
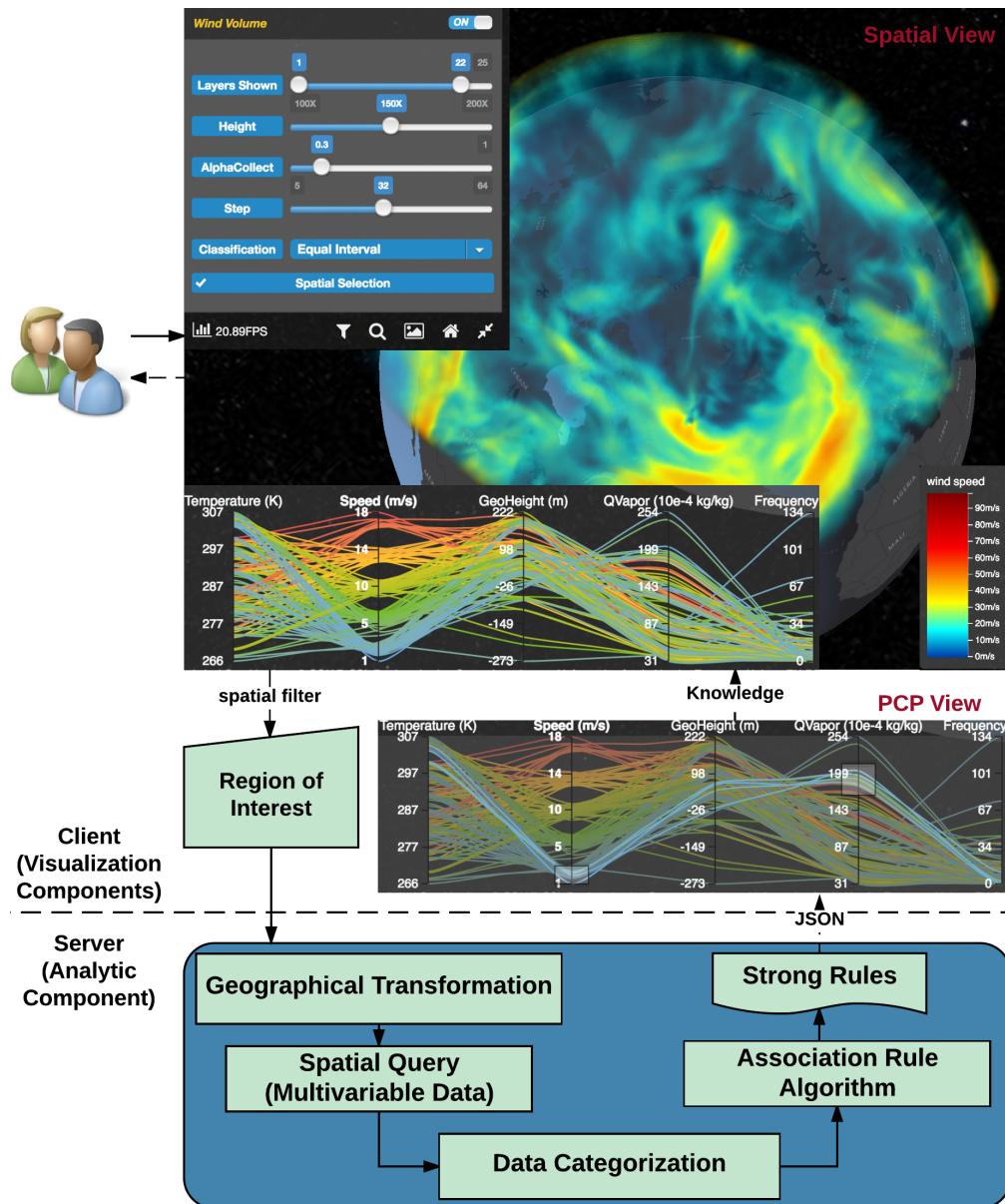
In this study, the Arctic System Reanalysis (ASR) data [27], which have the characteristics listed above, are exploited for validating the effectiveness of the heuristic system in analyzing climate conditions in the Arctic region. The ASR data have a moderate spatial (30 km) and high temporal (3 h) resolution across the continental scale. The gridded output of ASR data has a cube size of $360 \times 360 \times 29$ (latitude, longitude, and altitude/pressure level) and spans from 2000 to 2012. Twelve variables, including air temperature, ice depth, humidity, etc., are simulated and recorded for reconstruction of the Arctic's climate variability and change.

4. Visual Analytics Techniques

Visualization and analytics are two fundamental components for data comprehension in a visual analytic system. In this section, we illustrate how both techniques are applied in our proposed heuristic system. According to Shneiderman [28], a general data exploration process is "overview first, zoom and filter, then detail-on-demand". Following this mantra, climate scientists visually filter the event or abnormality in the spatial view and derive knowledge from multiple variables when studying an atmospheric phenomenon. To support this task, the heuristic system visualizes the preselected variable in real time. The user's interactions, including zooming and filtering, then facilitate the discovery and identification of the phenomena of interest. Furthermore, the heuristic analytic components are employed to derive new knowledge or association rules among the variables.

Figure 1 depicts the workflow of data exploration and how the components in the system work together. The visualization component, implemented on the client end and accelerated by the graphic process unit (GPU), is comprised of the spatial view and the PCP view. We designed and implemented multiple widgets, including spatial filtering and multi-dimensional brushing, to support interaction. On the server end, the analytic component processes the filtered data and employs an association rules-learning algorithm to derive knowledge to deepen our understanding of polar cyclones.

Because Arctic cyclones significantly influence the warming and melting of the Arctic icecap [29], we illustrate the proposed visualization system with a polar cyclone and select the four variables most relevant to the formation and intensification of the cyclone to study the intervariable association: (1) air temperature (T), the temperature difference creates a pressure imbalance, which is the driving force for the formation of cyclone; (2) wind speed (S), the wind speed indicates the strength of the cyclone; (3) geopotential height (G), can be used for locating troughs and ridges; and (4) atmospheric water vapor (V), which releases heat when condensing in the atmosphere that helps to fuel the cyclone. Key techniques that support this knowledge-driven visualization pipeline are described in the following sections.
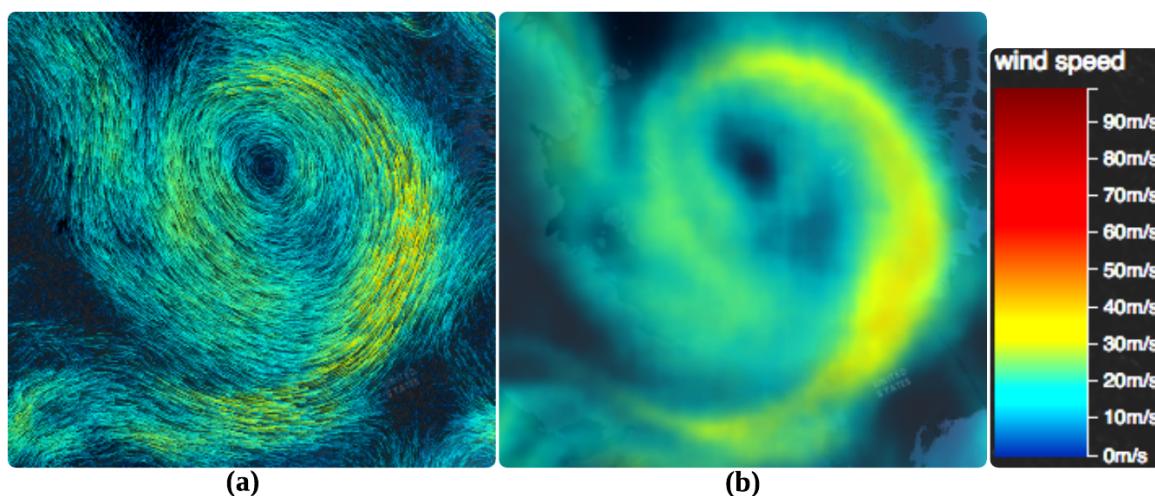
Figure 1. Workflow of association rule-based multivariate analysis and visualization. Multiple climate data processing services are provided on the server side. The multiview technique is employed to visualize data on the client side.

4.1. Spatiotemporal Data Visualization

Climate modeling scientists focus on the research of complicated climate physics and their compounding effects [30]. The ability to simulate atmospheric conditions over long time periods has led to a new understanding [31]. In this context, we select a time-varying wind field as the vector input of the simulation since it is an important indicator of extreme phenomena, such as polar cyclones. Although several previous studies have been conducted on vector data visualization, the pursuit of effective representation of time-varying vector flow remains an active research topic [32,33]. We select two visualization strategies, streamlines and volume rendering, to simulate the wind field data because streamlines yield the pattern of a cyclone, and volume rendering supports the internal data exploration in comparison to other existing techniques, such as particle tracking [29] and line integral convolution [34]. We accelerate the two approaches with GPUs to achieve better performance. Figure 2 depicts the visualization result of the Great Arctic Cyclone of 2012 with the two proposed approaches.

Streamline: Streamline algorithms have been widely used in flow visualization [35–37]. Achieved by dynamic seed distribution, the animated streamline-based flow provides an intuitive illustration of the vector field. A good example is Cameron Beccario's work [38] that focuses on developing a live visualization of global weather conditions. The resulting global wind visualization demonstrations have attracted significant attention. However, an issue with this method is that the response latency, or wait time, is high when the user interacts with the virtual earth or changes the viewpoint because the process of seed reinitialization and streamline regeneration is rather time-consuming. To improve visualization efficiency, we chose to reinitialize only the seed that is beyond the boundary, and keep the other parameters in their original positions. Given that the user's interaction is smooth and the camera view changes gradually in every frame, the seed's disturbance can be ignored. Visually, we achieve a smoother visualization than the existing work [38].

Figure 2. Screenshot of cyclone visualization with different visualization techniques: (**a**) streamline; (**b**) volume rendering.

Volume rendering: Climate scientists need to be able to explore the internal patterns of the data because a cyclone is an atmospheric phenomenon hovering in 4D space (3D + Time). Volume rendering, which is widely used in scientific visualization, is an effective technique to explore the data inside a volume [39]. In this work, we employ a GPU-based volume ray casting algorithm to gain insight into the wind field. An introduction to the algorithm can be found in [40,41]. By highlighting the high-value region of a cyclone through opacity, the cyclone's skeleton emerges, as illustrated in Figure 2b. We calculate the opacity with the equation $opacity(voxel) = [Speed]^2$ to emphasize the high speed. To better understand the 4D volume data, we developed a user-defined color scheme widget that allows the user to update the color transfer function in real time. The rendering result is updated with the movement of the control point in the color bar.
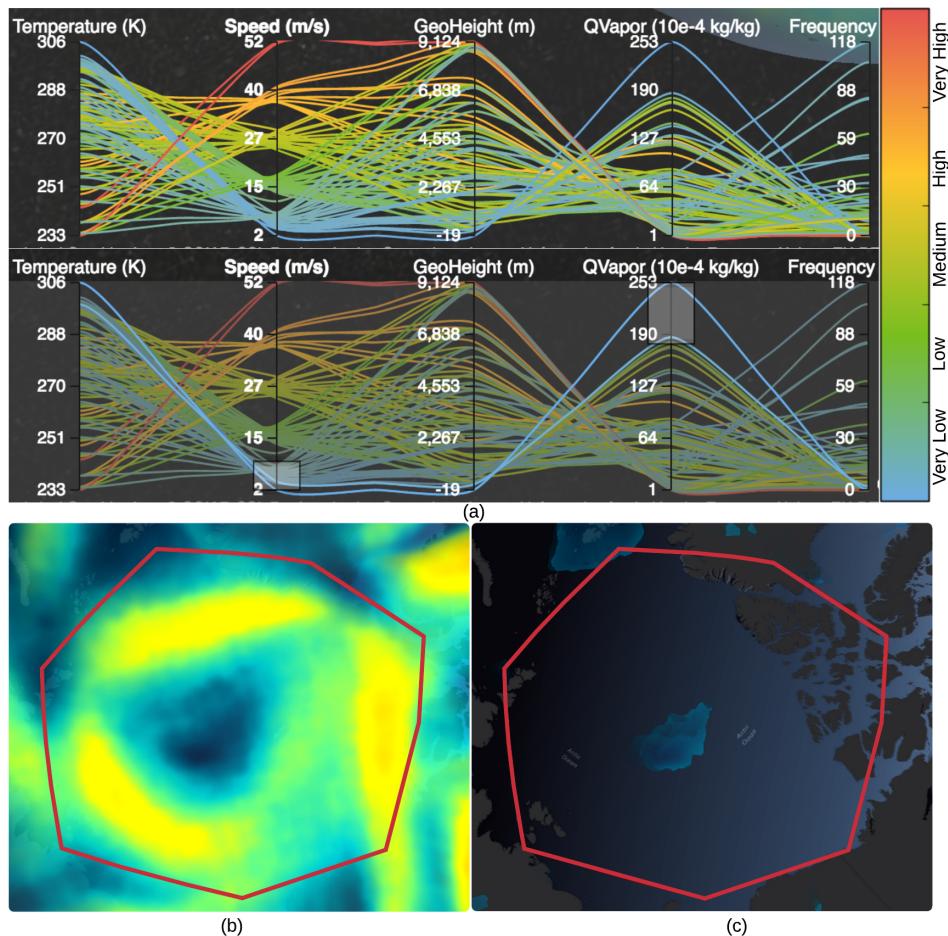
4.2. Spatial Filter

The spatial coverage of the demonstrated climate data ranges from 180 W to 180 E in longitude and 24.716 N to 90 N in latitude. Analyzing and mining the original data within the entire area is complex and confusing, especially when a scientist wants to focus on a local study area. Hence, users need an interactive spatial filtering tool to explore a region of interest.

Three coordinate systems are involved in the spatial filtering process: a screen coordinate system, a spherical coordinate system, and a model coordinate system. The spatial filter enables users to custom draw a region of interest on the virtual globe. Hence, geographical transformation must precede data filtering. The user's input coordinate (X, Y) in the screen coordinate system is transformed to a WGS84 coordinate $(Longitude, Latitude)$ in the spherical coordinate system on the globe, and then

mapped to the grid coordinate (X, Y, Z) in the model coordinate system (see [42] for the fundamental transformation formulations). The spatial filter takes place via two steps: (1) transform the user-defined bounding box from the screen coordinate to the grid coordinate and (2) conduct a spatial query in the grid coordinate by checking if a voxel is within the bounding box. The grid points falling outside the selected study area are excluded from visualization or analysis in the later phases.

4.3. Multiview Data Exploration

After capturing an overview of the climate data from the spatial view, the system then engages a multifaced PCP view to assist multivariate analysis. PCP, a technique applied to a diverse set of multi-dimensional problems, provides a more vivid output and a friendlier interface than other multivariate analysis techniques, such as scatterplot [43]. PCP plots multiple variables observed at the same spatial coordinates with a connected line. Qualitative relations emerge after all data have been plotted. Figure 3a illustrates the associations of four variables observed near a cyclone center. The top image shows all the connections, with the color scheme applied to the variable "wind speed". The bottom image highlights the low-speed lines. Results show that observation points with a relatively low speed are usually associated with a very high temperature and high water vapor near a cyclone center.

Figure 3. Illustration of the interaction of the spatial view and parallel coordinate plot (PCP) view: (**a**) multifaced PCP view; (**b**) user interacts in the spatial view; (**c**) updating of the spatial view along with PCP.

To support the second stage of Shneiderman's mantra, we developed two useful filter approaches: (1) a spatial filter in the spatial view to bound the region of interest (ROI) and (2) a multi-dimensional

data filter in the multifaced PCP view through brushing. Once a polar cyclone is detected in the spatial view, users can bound the cyclone with an editable bounding box (red polygon in Figure 3b). The data in the PCP view are updated collaboratively to display the selected volume grid within ROI. Meanwhile, brushing the data in the PCP view enables the user to filter associations of interest and trigger the spatial view filter as well. In Figure 3a, when the user brushes the axis of wind speed and water vapor to filter the grid with very low wind speed and high water vapor (humidity), the cyclone eye instantly emerges within the user-defined ROI (Figure 3c).

One drawback of the PCP view is the visual clutter issue when a plethora of lines are visualized [30]. In our demonstration, there are 3,240,000 voxels in the ASR data. Drawing so many lines in a real-time visualization system leads to a barely recognizable PCP view and limits interaction. Rather than allocating one line for each voxel, the variables are first categorized into several classes (see categorization details in Section 4.5), and then lines falling in the same category are grouped together and represented by a single line. The line's placement on each axis will be an averaged value of all lines sharing similar trends. The frequency of each line (unit: ‰) is calculated and displayed on the fifth axis. Another improvement for the PCP view is colorization. By clicking the axis title, all displayed lines will be recolored according to the value distribution of selected variables. See the effect in Figure 3a when a color scheme is applied on the variable "wind speed".

4.4. Mining Algorithm

Besides the visualization component, another main contribution of our work is the knowledge-driven component, which integrates an association rule-learning algorithm with multivariate analysis. Association rule learning [44], used to discover interesting relations between variables in large databases, was initially used for market basket analysis. It has gradually been extended to other application domains, such as bioinformatics, earth science, and scientific data analysis. An example in the area of earth science is the interesting connection revealed by association patterns that show how the different elements of the earth system interact with each other [10].

An association rule is indicated by the form $X \rightarrow Y$, which states that a random user who selects itemset X is also likely to select itemset Y. An itemset is a collection of zero or more items in a commodity pool $I = \{i_1, i_2, ..., i_k\}$. An important property of an itemset is its support count, which refers to the number of transactions contained in a specific itemset. For itemset X, the support count $\lambda(X)$ in a transaction database $T = \{t_1, t_2, ..., t_n\}$ (t_i is the ith transaction record) can be stated as follows:

$$\lambda(X) = count(\{t_i | X \subseteq t_i, t_i \in T\}). \tag{1}$$

The estimated criteria of a rule usually include the terms of support and confidence. Support is an indication of the probability of the antecedent, represented by $P(X)$. Confidence determines how frequently item Y appears in transactions that contain X, represented by $P(Y|X)$. The formal definitions of these metrics are:

$$Support, P(X) = \frac{\lambda(X \cup Y)}{N}, \tag{2}$$

$$Confidence, P(Y|X) = \frac{\lambda(X \cup Y)}{\lambda(X)}. \tag{3}$$

The goal of association rule learning is to find rules of interest from a large transaction database T whose support and confidence are larger than the corresponding thresholds. The thresholds ranging from 0 to 1 are adjustable according to the specific application in our heuristic system. To analyze the association of climate data, we make an analogy of climate data to transaction records. All voxels in the time-varying volume data constitute a transaction database. Categorized variables (see details in Section 4.5) in each voxel can be a selected item in the commodity pool, for example, item pool I = {VLT,

LT, MT, HT, VHT, VLS, LS, MS, HS, VHS} when temperature and wind speed are selected to conduct the mining process. Itemsets X (i.e., VLS) and Y (i.e., HT) denote an arbitrary combination of items in I, with the restriction that only one of the five categories of a variable can be selected to put in X or Y. To derive the association rules, we first calculate the support and confidence using Equations (2) and (3). A rule, e.g., $(VLS) \rightarrow (HT)$, is determined to be a strong rule if the possibility is higher than the pre-set minimum support and confidence values. Usually, the minimum confidence should be larger than the minimum support. The larger the threshold of minimum support and confidence, the stronger the rule is.

In this work, we implemented a classic association rule algorithm—the a priori algorithm [44]—to discover the hidden associations among the variables. The variable values of all spatially filtered points are categorized (see details in Section 4.5) and then dumped into the a priori model for rule learning. The generation process is usually split into two steps: (1) find all the itemsets whose support is larger than the minimum support threshold (the selected itemsets are called frequent itemsets) and (2) apply the minimum confidence constraint to the frequent itemsets and form the strong rules.
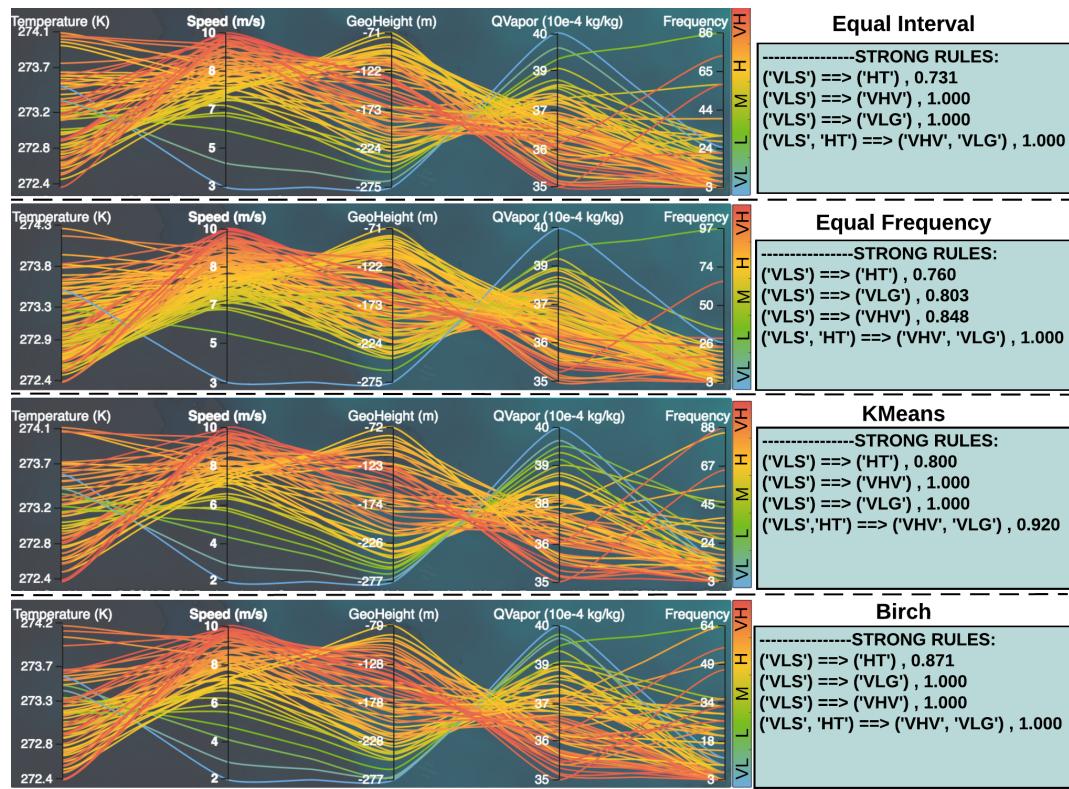
Finding the frequent itemsets is a time-consuming task since it involves searching all possible itemsets. The size of the itemsets grows exponentially, which makes the query difficult since the number of voxels is up to 3,240,000. However, by applying the downward-closure property [45], which states that an arbitrary subset of a frequent itemset is also frequent, the efficiency of the rule generation can be substantially improved.

4.5. Data Categorization

Variables in most scientific datasets are continuous values. Data categorization has thus become an important step. In our work, data categorization benefits both association rule mining and PCP generation. The association rule-mining algorithm is more efficient for processing categorical and discrete variables, although it was originally designed to deal with datasets containing Boolean-type attributes. Meanwhile, superabundant connected lines in the PCP view will cause a visual clutter issue and increase interaction latency if there are many voxels.

In this work, each variable's value domain is discretized into five categories: very high (VH), high (H), medium (M), low (L), and very low (VL). Combining the category label with the variable label, we can generate the abbreviation for each categorized variable (T, S, G, V). For instance, air temperature data (T) can be represented using VHT, HT, MT, LT, and VLT. To investigate the effect of different categorization methods in rule mining and visualization, we apply four popular unsupervised strategies to the raw data: equal interval, equal frequency, K-means [46], and birch clustering [47].
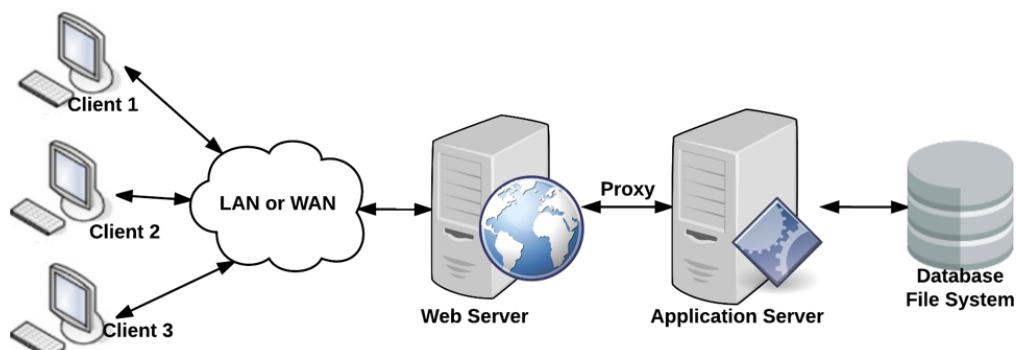
The data in the bottom vertical layer (at 100 hPa) are tested to reveal the underlying pattern because a cyclone is a near-surface atmospheric phenomenon [21]. Figure 4 depicts the outcome of the PCP view and strong association rules when applying these different strategies to the filtering data within a great cyclone near Alaska on 08/06/12. The line distribution and relations share similar patterns in the four PCP views. Taking temperature (T) and wind speed (S) as an example, the low-speed voxels are normally associated with a high temperature, but the high-speed voxels do not show strong relations with temperature. Likewise, the generated strong rules show that different strategies have only a small influence on the mining result. Although the strength of the first rule ($VLS \rightarrow HT$) varies somewhat when categorization methods are different, all generated rules remain the same. This experiment shows that the heuristic system is robust towards data classification algorithms. Eventually, we selected K-means in our heuristic system after balancing the algorithm's performance and efficiency. Although more simple and time efficient, equal interval and equal frequency methods can result in bad categorization results when a variable is distributed unevenly within its value range. On the other hand, birch is more time-consuming than K-means since it maintains a clustering feature tree.

Figure 4. Data visualization and analysis outcome using four different category strategies. Top to bottom: equal interval, equal frequency, *K*-means and birch.

5. System Implementation

Figure 5 illustrates the system architecture of the proposed heuristic climate visualization system. On the server side, two distributed servers (a web server and an application server) are responsible for conducting the knowledge-driven process. The web server responds to the user's resource request (HTML, image, climate attributes, etc.) as well as forwards the calculative request to the application server through a web proxy. The application server implemented in Python is mainly responsible for (1) accessing climate data; (2) the spatial filter according to the user-defined boundary; (3) data categorization; and (4) the a priori mining process. This multiscale server architecture increases the maintainability and extensibility of the system. It is simple to maintain and upgrade, highly expandable, and adaptable to increased workloads.

Figure 5. Browser/server architecture of our heuristic system. An application server is used to process the climate data. The web server communicates with an application server through a web proxy.

At the client end, the browser visualizes the time-varying climate data on a virtual globe in real time [48]. JavaScript and D3 library were employed to implement and tessellate (tile) the spatial view and PCP view into the Cesium platform [49] (an open-source JavaScript library for world-class globes based on WebGL). Users bound a research area and trigger a mining request after they discover a phenomenon of interest. The request is posted to the server through the Internet. The heuristic system also allows users to select an arbitrary combination of variables to generate the strong association rules with user-defined minimum support and confidence. This way, a user can flexibly select important variables for analyzing different atmospheric phenomena.
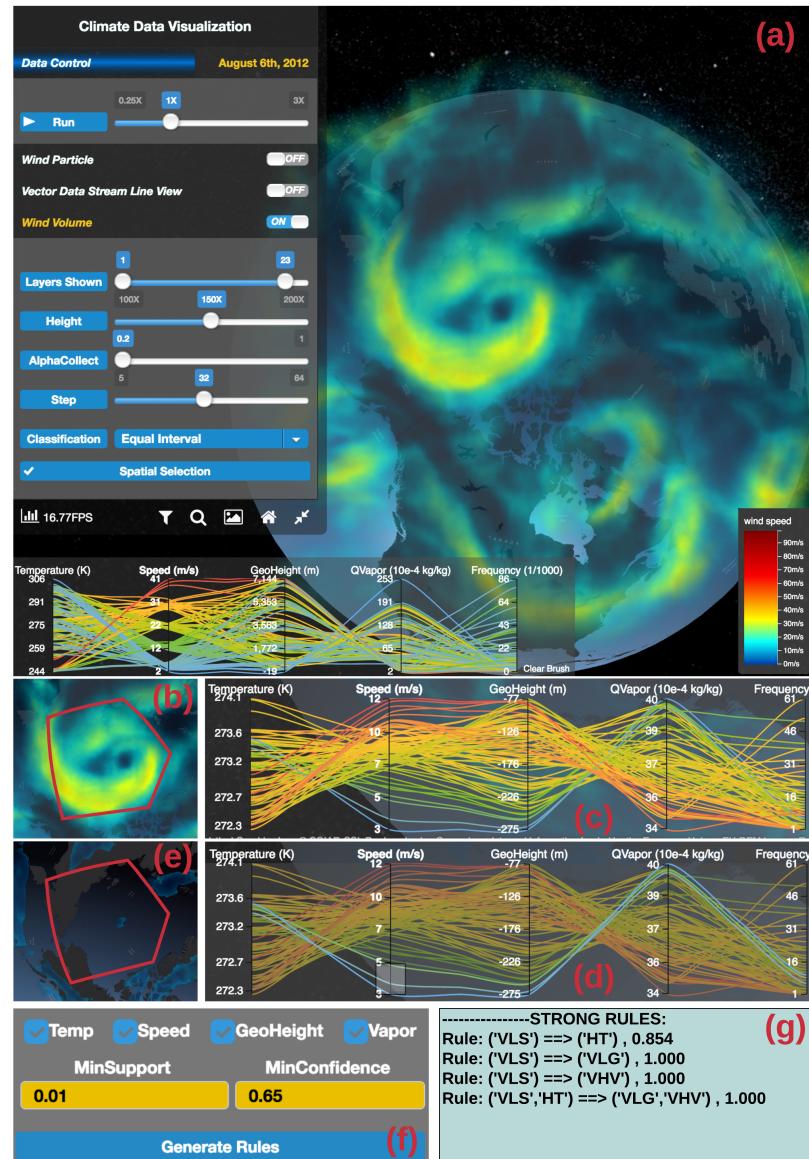
6. Case Study: The Great Arctic Cyclone of 2012

In this section, a famous cyclone named "the Great Arctic Cyclone of 2012" is used to demonstrate the effectiveness of our heuristic system. This cyclone was an unusually strong storm that formed off the coast of Alaska on 6 August 2012, and tracked into the center of the Arctic Ocean, where it slowly dissipated. According to the records, it was the strongest summer storm and the 13th strongest storm observed at any time in the Arctic since satellite observation began in 1979 [50,51]. The demonstration of our system was run on a desktop machine with a 2.7 GHz Intel Core i5 CPU (Tempe, AZ, USA), 16 GB of RAM, and an AMD Radeon HD 6770m GPU (Tempe, AZ, USA) with 512 MB texture memory. Figure 6 shows a user case diagram depicting the workflow of climate data exploration. Following the diagram labels from (a) to (g), climate scientists obtain a qualitative and quantitative analysis of the polar cyclone.
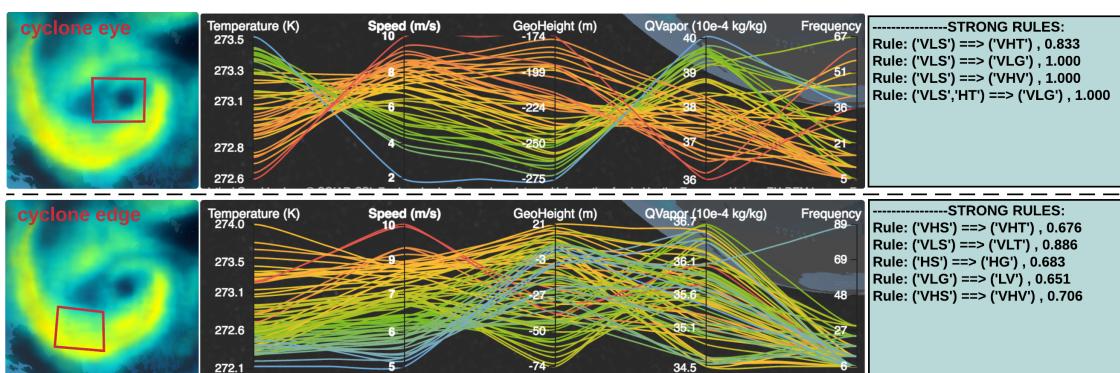
Within the research region, the perfect shape of the study cyclone is discovered and tracked in the spatial view after we set the timeline to 6 August 2012 (Figure 6a). The PCP view demonstrates the data distribution in the research region. To further explore the cyclone, a user-defined spatial filter is conducted near the cyclone eye (Figure 6b), which triggers the update of data in the PCP view (Figure 6c). The eye of a strong cyclone is characterized by light wind speed [52]. This characteristic is exploited to filter out the cyclone eye. The wind speed ranges from 0 to 12 m/s in the selected region. We then brush the data ($speed = VLS$, range from 0 to 4 m/s) in the PCP view. The association of the four variables relative to the cyclone eye is highlighted in the PCP view in Figure 6d, and the eye emerges in the spatial view in Figure 6e. The highlighted associations indicate that the cyclone eye wall is labeled with high temperature (HT, 273.3∼273.5 F), low geopotential height (VLG, $-273 \sim -261$ m), and dense water vapor (VHV, $39.3 \sim 40 \times 10e - 4$ kg/kg). A widget is also developed for calculating the strength of the association rules and identifying strong associations (see illustration in Figure 6f). In the demonstration, association rules are calculated with a user-defined minimum support of 0.01 and minimum confidence of 0.65. The higher value indicates the stronger rules that are derived. By filtering the generated rules with additional criteria (i.e., set wind speed to be VLS), the strong rules near the cyclone eye are picked out and depicted in Figure 6g.

Nine powerful cyclones in 2012 are selected to analyze the association. Table 1 illustrates the strong association rules generated near each cyclone eye. The demonstration shows that the polar cyclone eye wall presents with minimum surface pressure and large water vapor, which yields heavy snow in the North Pole. The association between temperature and speed near the cyclone eye is more complicated. In the polar region ($latitude \geq 66.5$ N), the cyclone eye is more likely to have a high temperature (warm core), whereas in the extratropical region (23.5 N $\leq$ latitude $\leq$ 66.5 N), the cyclone eye is characterized by a low temperature (cold core).

A comparison of data and rules between the cyclone eye and cyclone edge is illustrated in Figure 7, in which different data associations in the PCP view are captured. The generated rules show that data near the cyclone eye are more likely to have a negative correlation, whereas data near cyclone edge have a positive correlation. Low speed has a strong association with the high temperature near the cyclone eye, but it is just the opposite near the cyclone edge. The results indicate that this association might be used as a criterion of cyclone identification.

Figure 6. A user case study of climate dataset exploration in our system. The cyclone tracked is "The Great Arctic Cyclone of 2012" on 6 August 2012.

Figure 7. Comparison of data distribution pattern in the PCP view and generated rules between cyclone eye and cyclone edge. The data distribution patterns in the PCP view are different, leading to the discovery of different strong association rules.

Table 1. Strong association rules that are generated near the cyclone eye in 2012; VLS, LS, MS, HS, and VHS represent very low, low, medium, high, and very high speed, respectively

Timestamps	Latitude/Longitude	Strong Rules
1 January 2012	89.45 N/6.44 W	$(VLS)\rightarrow(HT)$, 0.833 $(VLS)\rightarrow(VLG)$, 1.000 $(VLS)\rightarrow(HV)$, 0.767
12 January 2012	47.94 N/162.06 E	$(VLS)\rightarrow(LT)$, 0.625 $(VLS)\rightarrow(VLG)$, 1.000 $(VLS)\rightarrow(VHV)$, 0.825
15 April 2012	51.32 N/157.23 W	$(VLS)\rightarrow(LT)$, 0.875 $(VLS)\rightarrow(VLG)$, 1.000 $(VLS)\rightarrow(HV)$, 0.873
14 May 2012	65.91 N/1.86 W	$(VLS)\rightarrow(MT)$, 0.750 $(VLS)\rightarrow(VLG)$, 1.000 $(VLS)\rightarrow(HV)$, 0.893
14 May 2012	72.28 N/86.09 W	$(VLS)\rightarrow(HT)$, 0.706 $(VLS)\rightarrow(VLG)$, 1.000 $(VLS)\rightarrow(HV)$, 0.893
31 May 2012	48.94 N/34.82 W	$(VLS)\rightarrow(LT)$, 0.778 $(VLS)\rightarrow(VLG)$, 1.000 $(VLS)\rightarrow(HV)$, 0.893
7 August /2012	83.03 N/178.24 W	$(VLS)\rightarrow(HT)$, 1.000 $(VLS)\rightarrow(VLG)$, 1.000 $(VLS)\rightarrow(VHV)$, 1.000
21 October 2012	48.46 N/31.22 W	$(VLS)\rightarrow(LT)$, 1.000 $(VLS)\rightarrow(VLG)$, 1.000 $(VLS)\rightarrow(VHV)$, 0.800
3 November 2012	53.30 N/153.23 W	$(VLS)\rightarrow(MT)$, 1.000 $(VLS)\rightarrow(VLG)$, 1.000 $(VLS)\rightarrow(VHV)$, 0.857

7. Conclusions

In this paper, we propose a web-based heuristic climate visualization system to explore spatiotemporal multivariate data. Real-time interactive climate data visualization and association rules mining capabilities are developed that allow for interactive vector field streamline and volume rendering visualization. In the visualization process using streamlines, an improvement for better interaction is achieved by reinitializing only the portion of the streamline seed outside the boundary. To discover underlying knowledge in the data, association rules learning is used to mine strong association rules between multiple variables. Additionally, multiple data exploratory methods (e.g., spatial filter, multifaced PCP view, brushing, and linking) are employed to facilitate data exploration.

We then integrate the submodules and multiple visualization techniques to explore and analyze time-varying climate data for the Great Arctic Cyclone of 2012 and test the system's effectiveness and usefulness. Results show that our system provides an effective tool for users to explore climate data. For example, some general principles about the cyclone eye wall were derived and analyzed from the data, which is useful for identifying cyclones.

Currently, only one variable at a time is visualized in the spatial view within the proposed visualization framework. In the future, we will add interactive visualization of multiple variables in the spatial view. To further develop this complex heuristic visualization system and integrate multiple heterogeneous data sources seamlessly, more research needs to be conducted on analyzing multivariable correlation with other data mining algorithms, such as neutral network and deep learning.

References

1. Li, Z.; Yang, C.; Sun, M.; Li, J.; Xu, C.; Huang, Q.; Liu, K. A High Performance Web-Based System for Analyzing and Visualizing Spatiotemporal Data for Climate Studies. In *Web and Wireless Geographical Information Systems*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 190–198.
2. Akiba, H.; liu Ma, K.; Chen, J.; Hawkes, E. Visualizing Multivariate Volume Data from Turbulent Combustion Simulations. *Comput. Sci. Eng.* **2007**, *9*, 76–83. [CrossRef]

3.   Nocke, T.; Sterzel, T.; Böttinger, M.; Wrobel, M. Visualization of climate and climate change data: An overview. In *Digital Earth Summit on Geoinformatics 2008: Tools for Global Change Research (ISDE'08)*; Wichmann: Heidelberg, Germany, 2008; pp. 226–232.

4.   Biswas, A.; Dutta, S.; Shen, H.W.; Woodring, J. An Information-Aware Framework for Exploring Multivariate Data Sets. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2683–2692. [CrossRef] [PubMed]

5.   Turkay, C.; Filzmoser, P.; Hauser, H. Brushing Dimensions—A Dual Visual Analysis Model for High-Dimensional Data. *IEEE Trans. Vis. Comput. Graph.* **2011**, *17*, 2591–2599. [CrossRef] [PubMed]

6.   Guo, H.; Xiao, H.; Yuan, X. Scalable multivariate volume visualization and analysis based on dimension projection and parallel coordinates. *IEEE Trans. Vis. Comput. Graph.* **2012**, *18*, 1397–1410. [PubMed]

7.   Arthur, D.K.; Lasher-Trapp, S.; Abdel-Haleem, A.; Klosterman, N.; Ebert, D.S. A New Three-Dimensional Visualization System for Combining Aircraft and Radar Data and Its Application to RICO Observations. *J. Atmos. Ocean. Technol.* **2010**, *27*, 811–828. [CrossRef]

8.   Ayachit, U. *The Paraview Guide: A Parallel Visualization Application*; Kitware, Inc.: Clifton Park, NY, USA, 2015.

9.   Williams, D.N.; Bremer, T.; Doutriaux, C.; Patchett, J.; Williams, S.; Shipman, G.; Miller, R.; Pugmire, D.R.; Smith, B.; Steed, C.; et al. Ultrascale visualization of climate data. *Computer* **2013**, *46*, 68–76. [CrossRef]

10.   Liu, X.; Shen, H.W. Association Analysis for Visual Exploration of Multivariate Scientific Data Sets. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 955–964. [CrossRef] [PubMed]

11.   Kehrer, J.; Hauser, H. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 495–513. [CrossRef] [PubMed]

12.   Ding, Z.; Ding, Z.; Chen, W.; Chen, H.; Tao, Y.; Li, X.; Chen, W. Visual inspection of multivariate volume data based on multi-class noise sampling. *Vis. Comput.* **2015**, *32*, 465–478. [CrossRef]

13.   Woodring, J.; Shen, H.W. Multi-variate, time varying, and comparative visualization with contextual cues. *IEEE Trans. Vis. Comput. Graph.* **2006**, *12*, 909–916. [CrossRef] [PubMed]

14.   Wang, C.; Yu, H.; Ma, K.L. Importance-Driven Time-Varying Data Visualization. *IEEE Trans. Vis. Comput. Graph.* **2008**, *14*, 1547–1554. [CrossRef] [PubMed]

15.   Jankun-Kelly, T.J.; Ma, K.L. A Study of Transfer Function Generation for Time-Varying Volume Data. In *Eurographics*; Springer: Vienna, Austria, 2001; pp. 51–65.

16.   Jänicke, H.; Böttinger, M.; Scheuermann, G. Brushing of attribute clouds for the visualization of multivariate data. *IEEE Trans. Vis. Comput. Graph.* **2008**, *14*, 1459–1466. [CrossRef] [PubMed]

17.   Potter, K.; Wilson, A.; Bremer, P.T.; Williams, D.; Doutriaux, C.; Pascucci, V.; Johnson, C.R. Ensemble-vis: A framework for the statistical visualization of ensemble data. In Proceedings of the ICDMW'09. IEEE International Conference on Data Mining Workshops, Miami, FL, USA, 6–9 December 2009; pp. 233–240.

18.   Luo, S.; Dingliana, J. Selective Saturation and Brightness for Visualizing Time Varying Volume Data. In Proceedings of the EG/VGTC Conference on Visualization (EuroVis) 2015 Posters, Clagary, Italy, 25–29 May 2015.

19.   Wang, C.; Yu, H.; Grout, R.W.; Ma, K.L.; Chen, J.H. Analyzing information transfer in time-varying multivariate data. In Proceedings of the 2011 IEEE Pacific Visualization Symposium. Institute of Electrical and Electronics Engineers (IEEE), Hong Kong, China, 1–4 March 2011, doi:10.1109/pacificvis.2011.5742378. [CrossRef]

20.   Gosink, L.; Anderson, J.; Bethel, W.; Joy, K. Variable interactions in query-driven visualization. *IEEE Trans. Vis. Comput. Graph.* **2007**, *13*, 1400–1407. [CrossRef] [PubMed]

21.   Li, W.; Yang, C.; Sun, D. Mining geophysical parameters through decision-tree analysis to determine correlation with tropical cyclone development. *Comput. Geosci.* **2009**, *35*, 309–316. [CrossRef]

22.   Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2815–2831, doi:10.5194/nhess-13-2815-2013. [CrossRef]

23.   Mounce, S.; Husband, S.; Furnass, W.; Boxall, J. Multivariate Data Mining for Estimating the Rate of Discoloration Material Accumulation in Drinking Water Systems. *Procedia Eng.* **2014**, *89*, 173–180. [CrossRef]

24.   Yang, R.; Sun, D.; Tang, J. A "sufficient" condition combination for rapid intensifications of tropical cyclones. *Geophys. Res. Lett.* **2008**, *35*. [CrossRef]

25.   Yang, R.; Tang, J.; Sun, D. Association rule data mining applications for Atlantic tropical cyclone intensity changes. *Weather Forecast.* **2011**, *26*, 337–353. [CrossRef]

26. Faghmous, J.H.; Kumar, V. Spatio-temporal Data Mining for Climate Data: Advances, Challenges, and Opportunities. In *Studies in Big Data*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 83–116.

27. Bromwich, D.; Kuo, Y.H.; Serreze, M.; Walsh, J.; Bai, L.S.; Barlage, M.; Hines, K.; Slater, A. Arctic system reanalysis: Call for community involvement. *Eos Trans. Am. Geophys. Union* **2010**, *91*, 13–14. [CrossRef]

28. Ferster, B.; Shneiderman, B. *Interactive Visualization: Insight through Inquiry*; MIT Press: Cambridge, MA, USA, 2012.

29. Wang, F.; Li, W.; Wang, S. Polar Cyclone Identification from 4D Climate Data in a Knowledge-Driven Visualization System. *Climate* **2016**, *4*, 43. [CrossRef]

30. Wong, P.C.; Shen, H.W.; Leung, R.; Hagos, S.; Lee, T.Y.; Tong, X.; Lu, K. Visual analytics of large-scale climate model data. In Proceedings of the 2014 IEEE 4th Symposium on Large Data Analysis and Visualization (LDAV), Paris, France, 9–10 November 2014. [CrossRef]

31. Goddard, P.B.; Yin, J.; Griffies, S.M.; Zhang, S. An extreme event of sea-level rise along the Northeast coast of North America in 2009–2010. *Nat. Commun.* **2015**, *6*, 6346. [CrossRef] [PubMed]

32. Johnson, C. Top scientific visualization research problems. *IEEE Comput. Graph. Appl.* **2004**, *24*, 13–17. [CrossRef] [PubMed]

33. Fuchs, R.; Hauser, H. Visualization of Multi-Variate Scientific Data. *Computer Graphics Forum*; Wiley Online Library: Hoboken, NJ, USA, 2009; Volume 28, pp. 1670–1690.

34. Sundquist, A. Dynamic line integral convolution for visualizing streamline evolution. *IEEE Trans. Vis. Comput. Graph.* **2003**, *9*, 273–282. [CrossRef]

35. Turk, G.; Banks, D. Image-guided streamline placement. In Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques-SIGGRAPH, New Orleans, LA, USA, 4–9 August 1996. [CrossRef]

36. Chen, C.K.; Yan, S.; Yu, H.; Max, N.; Ma, K.L. An Illustrative Visualization Framework for 3D Vector Fields. *Comput. Graph. Forum* **2011**, *30*, 1941–1951. [CrossRef]

37. Yu, H.; Wang, C.; Shene, C.K.; Chen, J.H. Hierarchical Streamline Bundles. *IEEE Trans. Vis. Comput. Graph.* **2012**, *18*, 1353–1367. [CrossRef] [PubMed]

38. Beccario, C. A Visualization of Global Weather Conditions Forecast By Supercomputers. Available online: https://earth.nullschool.net/ (accessed on 9 December 2016).

39. Zhang, L.; Wang, K.; Zuo, W. Real-time multi-volume rendering for 3d electrophysiological data visualization based on graphics processing unit. *ICIC Express Lett. Part B Appl. Int. J. Res. Surv.* **2013**, *4*, 1625–1630.

40. Callahan, S.P.; Callahan, J.H.; Scheidegger, C.E.; Silva, C.T. Direct volume rendering: A 3D plotting technique for scientific data. *Comput. Sci. Eng.* **2008**, *10*, 88–92. [CrossRef]

41. Feng, W.; Gang, W.; Deji, P.; Yuan, L.; Liuzhong, Y.; Hongbo, W. A parallel algorithm for viewshed analysis in three-dimensional Digital Earth. *Comput. Geosci.* **2015**, *75*, 57–65. [CrossRef]

42. Liu, P.; Gong, J.; Yu, M. Visualizing and analyzing dynamic meteorological data with virtual globes: A case study of tropical cyclones. *Environ. Model. Softw.* **2015**, *64*, 80–93. [CrossRef]

43. Inselberg, A.; Dimsdale, B. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In Proceedings of the First IEEE Conference on Visualization: Visualization, San Francisco, CA, USA, 23–26 October 1990. [CrossRef]

44. Agrawal, R.; Imieliński, T.; Swami, A. Mining association rules between sets of items in large databases. *ACM SIGMOD Rec.* **1993**, *22*, 207–216. [CrossRef]

45. Agrawal, R.; Srikant, R. Fast algorithms for mining association rules. In Proceedings of the 20th International Conference very Large Data Bases, VLDB, Santiago, Chile, 12–15 September 1994; Volume 1215, pp. 487–499.

46. Wagstaff, K.; Cardie, C.; Rogers, S.; Schrödl, S. Constrained k-means clustering with background knowledge. *ICML* **2001**, *1*, 577–584.

47. Lorbeer, B.; Kosareva, A.; Deva, B.; Softić, D.; Ruppel, P.; Küpper, A. A-BIRCH: Automatic Threshold Estimation for the BIRCH Clustering Algorithm. In Proceedings of the INNS Conference on Big Data, Thessaloniki, Greece, 23–25 October 2016; pp. 169–178.

48. Li, W.; Wang, S. PolarGlobe: A web-wide virtual globe system for visualizing multi-dimensional, time-varying, big climate data. *Int. J. Geogr. Inf. Sci.* **2017**, *2017*, 1–21. [CrossRef]

49. Keysers, J. *Review of Digital Globes 2015*; CRCSI: Victoria, Australia, 2015.

50. Parkinson, C.L.; Comiso, J.C. On the 2012 record low Arctic sea ice cover: Combined impact of preconditioning and an August storm. *Geophys. Res. Lett.* **2013**, *40*, 1356–1361. [CrossRef]

51. Simmonds, I.; Rudeva, I. The great Arctic cyclone of August 2012. *Geophys. Res. Lett.* **2012**, *39*. [CrossRef]
52. Zappa, G.; Shaffrey, L.; Hodges, K. Can Polar Lows be Objectively Identified and Tracked in the ECMWF Operational Analysis and the ERA-Interim Reanalysis? *Mon. Weather Rev.* **2014**, *142*, 2596–2608. [CrossRef]