

EVis: Visually Analyzing Environmentally Driven Events

Tinghao Feng, Jing Yang, Martha-Cary Eppes, Zhaocong Yang, Faye Moser

Abstract—Earth scientists are increasingly employing time series data with multiple dimensions and high temporal resolution to study the impacts of climate and environmental changes on Earth's atmosphere, biosphere, hydrosphere, and lithosphere. However, the large number of variables and varying time scales of antecedent conditions contributing to natural phenomena hinder scientists from completing more than the most basic analyses. In this paper, we present EVis (Environmental Visualization), a new visual analytics prototype to help scientists analyze and explore recurring environmental events (e.g. rock fracture, landslides, heat waves, floods) and their relationships with high dimensional time series of continuous numeric environmental variables, such as ambient temperature and precipitation. EVis provides coordinated scatterplots, heatmaps, histograms, and RadViz for foundational analyses. These features allow users to interactively examine relationships between events and one, two, three, or more environmental variables. EVis also provides a novel visual analytics approach to allowing users to discover temporally lagging relationships related to antecedent conditions between events and multiple variables, a critical task in Earth sciences. In particular, this latter approach projects multivariate time series onto trajectories in a 2D space using RadViz, and clusters the trajectories for temporal pattern discovery. Our case studies with rock cracking data and interviews with domain experts from a range of sub-disciplines within Earth sciences illustrate the extensive applicability and usefulness of EVis.

Index Terms—Multivariate Time Series, RadViz, Event Data, Visual Analytics, Earth Sciences

1 INTRODUCTION

Earth's environmental systems—its atmosphere, biosphere, hydrosphere, and lithosphere—are all linked together via complex pathways that make a prediction of changes in any one system extremely challenging. In particular, discrete ‘events’ such as rock fracture, flooding, and landslides, for example, are driven or limited by environmental factors like air temperature, rainfall, or river discharge. To best prepare for climate change's impacts, Earth scientists are monitoring **Environmental conditions (E-conditions)** at increasingly high temporal resolutions to understand **triggers**, namely the driving environmental factors, of natural event phenomena (hereafter events) [34, 37]. However, the large number of variables contributing to events, combined with the convolution of their covariance, has hindered a large majority of scientists from completing more than the most basic analyses [19]. Most researchers employ line plots, ‘wiggle matching’, Fourier, or bivariate analyses, and rely on hand filtering to test correlations within data subsets related to one or more dimensions. As a result, there is little room for the discovery of relationships not hypothesized by the viewers. Thus, there is a strong need for time series analysis tools that readily facilitate – without coding – more functionalities than plotting multicolored lines or heatmaps versus time.

Moreover, current approaches employed by Earth scientists lack the flexibility to readily data-mine for **temporally lagging relationships** like those related to antecedent conditions. The rates and magnitude of many environmental phenomena are predicated not only on simultaneous conditions but also on antecedent conditions for periods of minutes to days or weeks prior. Currently, the state of the science regarding lag-correlations requires scientists to know the period of the antecedence or to iteratively test multiple periods – a time-consuming and dissuading process. In addition, there is a need to readily explore the relationships among multiple environmental variables in the time(s) leading up to events of interest.

In this paper, we present EVis (Fig. 1), a fully working visual analytics prototype, developed to help scientists who have minimal coding

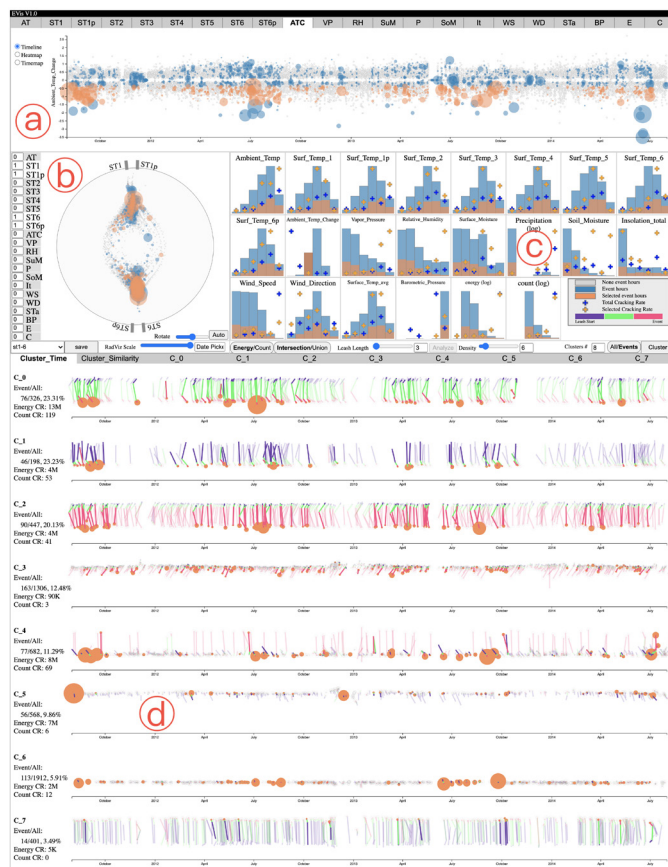


Fig. 1. The New Mexico Rock Dataset [17] in EVis. (a) The value-time scatterplot of Ambient Temperature Change (ATC). A group of discrete-time-points with negative ATCs are selected as POIs. (b) RadViz with two top surface temperature variables placed on the top and two bottom surface temperature variables placed at the bottom. (c) Histograms showing event hours and event rates of the whole dataset (in blue) and the POIs (in orange). (d) The cluster timeline view where three-hour leashes of the POIs are grouped into eight clusters and displayed over timelines. The clusters are sorted by event rate in descending order. Leashes of non-event hours are dimmed.

- Tinghao Feng, Jing Yang, and Zhaocong Yang are with Dept of Computer Science of UNC Charlotte. E-mail: tfeng1, jyang13, zyang19@unc.edu.
- Martha-Cary Eppes and Faye Moser are with Dept of Geography and Earth Sciences of UNC Charlotte. E-mail: meppes, fvisco@unc.edu.

or data mining expertise visually analyze time series of complex environmental data. EVIS was developed via a two-year collaboration between Earth scientists (scientists for short) and visualization researchers (vis-researchers for short). The New Mexico Rock dataset [16, 17], a previously well-studied dataset devoted to identifying key environmental variables driving natural rock cracking, was used to develop and evaluate the prototype. The scientists used sensors attached to natural rocks to capture rock cracking events and E-conditions in real-time. This dataset records 280,979 rock cracking events and 22 E-conditions simultaneously collected from a rock at an interval of one minute over three years. The overall goal of the research is to identify driving factors of the rock cracking events, with temporally lagging relationships related to antecedent conditions considered. Importantly, this dataset represents a typical exemplar of a broad array of scientific problems that may be addressed using continuous environmental monitoring in the context of any particular events of interest.

EVIS was created to allow scientists to evaluate known and discover unknown event triggers - not only in terms of simultaneous E-conditions when events happened, but also environmental changes and other events occurring before the events of interest occurred. Toward these goals, EVIS provides a set of basic and advanced graphical charts as well as a novel visual analytics approach, all coordinated in the same visual interface for a smooth visual analytics workflow. The basic charts include scatterplots, histograms, and heatmaps. The advanced chart is a RadViz view [25, 33] with new interaction techniques. They allow users to interactively explore relationships between events and one, two, three, or more simultaneous E-conditions. EVIS also provides a new visual analytics approach named **RadViz-Leash**. It integrates RadViz projection [25, 33], trajectory clustering, and visualizations to allow users to interactively explore temporally lagging relationships between events and multiple **Preceding Environmental conditions (PE-conditions)**, namely the environmental changes before the events occurred. The usefulness of EVIS is illustrated by two case studies conducted with a senior Earth scientist and feedback collected from 11 domain experts.

The main contributions of this paper include (1) EVIS, a visual analytics prototype specifically designed for Earth scientists to explore environmentally driven events and their relationships to multiple E-conditions, (2) RadViz-Leash, a new visual analytics approach for interactively analyzing temporally lagging relationships between events and their multidimensional preceding conditions, and (3) case studies and expert feedback to evaluate the usefulness of EVIS and RadViz-Leash.

2 BACKGROUND AND RELATED WORK

2.1 Rock Mechanical Weathering

One co-author of this paper is a senior Earth scientist. She and her team have been studying the type of rock mechanical weathering datasets presented herein for about 15 years. Mechanical weathering refers to the physical breakdown – cracking – of rock that naturally occurs when rocks are exposed to Earth’s atmosphere and to water. It is a key component of all Earth surface processes, and one for which data and understanding have been strongly lacking [15]. The scientists employed Acoustic Emission (AE) technology to study mechanical weathering by ‘listening’ to rock cracking events in real-time [16, 17]. They collected the dataset used in this paper from a granite boulder placed in Sevilleta National Wildlife Refuge in New Mexico [18]. An instrumentation system was affixed to the boulder, which consisted of multiple thermocouples, strain gauge rosettes, surface moisture sensors, and AE sensors. All sensors were installed on the boulder and calibrated in control. Measures of surface strain and temperature, surface moisture, and other E-conditions from an adjacent weather station were collected once a minute for about three years. During the same period of time, cracking events, as recorded by AE, were monitored continuously and recorded whenever they occurred. Each AE signal exceeding a defined threshold was recorded as an individual ‘event’ with associated ‘energy’ related to the strength of the captured signal. Therefore, in this case, cracking is the environmental event of interest, and that cracking also has an additional attribute of energy. This scenario is typical of other

Earth science applications. For example, the ‘event’ might be a flood, and the additional attribute volume of water.

Twenty-two E-conditions were recorded for 1,578,240 minutes. The E-conditions included not only sensor-collected variables, but also derived variables such as Vapor Pressure, which was calculated from Ambient Temperature and Relative Humidity. Meanwhile, 280,979 AE events were detected and recorded during this period. The dataset was fully vetted for quality in past work before this study.

In past work, the scientists employed stacked and annotated line charts (hand-selected and assembled in drawing software), histograms of hand-filtered data, and bivariate plots to analyze this dataset and similar datasets collected from other rocks [16]. Selecting days and hours of key data from the datasets required thousands of people-hours of data mining through hundreds of individual purpose-built graphs to find trends in a handful of the 22 dimensions available. They improved this approach by developing heatmaps of cracking rates under different E-conditions, but could only co-plot up to two dimensions simultaneously [17]. Based on the analyses presented in past publications, the scientists hypothesized not only that multiple dimensions are interconnected in contributing to observed cracking, but also that the temporal relationship between these dimensions and the onset of cracking varies. For example, they make the qualitative observation that cracking occurs in concert with rapid cooling immediately before cracking—brought on by both wind and rain—following week-long periods of similar hot days [16, 17]. To date, however, there has been no ability to readily test or explore these hypotheses. The scientists have identified a sequence of weather phenomena that appear to lead to events, but had not quantified that sequence in any way, nor identified repeated sequences of phenomena that lead to other events in the database.

2.2 Multivariate Time Series Visualization and Clustering

Many efforts have been made to visualize a collection of multivariate time series. Time searcher 2 [9] presented multivariate time series in a set of line charts, one for each dimension. Series selected in one line chart were highlighted in all line charts for interactive visual exploration. This approach was intuitive, but it did not explicitly reveal the temporal relationships among the dimensions. Fujiwara et al. [21] projected a group of multivariate time series to a 2D space as points using a two-step dimension reduction process. Users could examine feature contributions to clusters of time series in the projection space. Since each time series was reduced to one point in the projection space, this approach was not intuitive for users to compare temporal trends of a large number of time series in detail. Takami and Takama [41] used animations and trajectories to show how values of multiple objects change over time in a shared 2D space. Its drawbacks are the potential clutter and change blindness problems when a large number of objects were visualized.

Several methods have been proposed to visualize a single multivariate time series. Cheng et al. [11] divided a multivariate time series into time slices and projected them to an MDS plot based on their similarities. The temporal order of the time slices was mapped to the color of the projections. Time Curves [6] projected high dimensional temporal data points to a 2D space and connected them by their temporal order with a time curve. Characteristics of the time curve were extracted to reveal patterns of the temporally evolving data. Elzen et al. [43] projected snapshots of a dynamic network to dots in a 2D projection space and connected them using a curve. Visualizing a large number of time series was not a focus of these approaches.

The trajectory-based financial time series visualization [39] is the most relevant work of RadViz-Leash. It plotted 2D time series as trajectories in a 2D space and clustered the trajectories based on their high dimensional feature vectors using a Self-Organizing Map [27]. RadViz-Leash is also a trajectory and clustering based approach, but it is focused on multivariate time series and uses a different projection, clustering, and visualization approach.

Many algorithms have been developed to cluster multivariate time series. For example, Ghassempour et al. [23] clustered time series with both categorical and continuous variables based on hidden Markov models. Dasu and Swayne [13] clustered time series based on nonpara-

metric statistical summaries using a k-means algorithm. D'Urso and Maharaj [14] proposed wavelet-based clustering for multivariate time series. Singhal and Seborg [40] clustered multivariate time series using both a similarity factor based on principal component analysis and a distance similarity factor.

2.3 RadViz

RadViz [25,33] is a popular visualization technique for analyzing multidimensional data. It projects a multidimensional dataset to a 2D space where dimensions of the dataset are placed along the circumference of a circle as anchors. The projection of a data point is decided by anchor positions, value ranges used for normalizing the dimensions, and values of the data point via a spring force model. Springs connect the projection with the anchors. Their constants equal to normalized values of the data point on the anchor dimensions [25,33]. Cheng et al. [12] introduced RadViz Deluxe, which automatically relocated dimension anchors for clutter reduction and insight discovery. They employed several optimization procedures to enforce a variety of distance constraints. Angelini et al. [4] proposed a method to automatically rearrange dimension anchors based on a point disposition heuristic. Their prototype also allowed dragging the anchors freely along the circumference. Zhou et al. [47] provided rich interactions in RadViz, such as adding or removing dimensions, selecting data from RadViz, and examining them in a coordinated detail view. Their case studies revealed the benefits of providing interactions in RadViz.

3 DEVELOPMENT PROCESS AND REQUIREMENT ANALYSIS

The vis-researchers and scientists in our team have been working together on this project for two years. The first year was an exploration stage with bi-weekly meetings. The collaboration started with helping the vis-researchers understand the domain problem, the data, and the current practices in Earth sciences. Then, several Tableau [1] dashboards and ad hoc D3 [8] prototypes with different visualization designs were created using real datasets from the scientists. Some designs were suggested by the scientists based on their previous practices and leveraged by the vis-researchers. Others were contributed by the vis-researchers. The dashboards and prototypes were demonstrated to the scientists in face-to-face meetings for feedback.

The exploration stage was important for the collaboration – the scientists did not have a thorough understanding of what the vis-researchers could do for them at the beginning. New tasks and requirements often popped up after they were inspired by the visualizations brought to them. Also, the exploration allowed the vis-researchers to explore the design space and find the most desired visualizations by the scientists. For example, the RadViz-Leash was found to provide a ‘new’ view of the data, whereas an attempt to plot the daily time series of selected data was found to be redundant and unnecessary.

In the second year, EVis was developed using Node.js [42] and D3 [8]. It integrated the designs most positively perceived by the scientists. The team met weekly to report the development progress and collect feedback for recently developed functions. Throughout March 2021, the team met intensively (multiple one-hour Zoom meetings per week) to test the system using the pair analytics method [5]. In these meetings, a senior scientist explored her data using EVis. She drove the data exploration process and a vis-researcher conducted interactions for her. Usability issues noticed in the process were recorded and addressed before the next meeting. New functions were also added to EVis to address needs not captured in the original design but deemed desirable in the initial stages of evaluation. Section 5 recorded several case studies conducted in those meetings. EVis was also improved after receiving feedback from a group of domain experts (see Section 6).

Tasks and requirements for EVis became more and more clear when the collaboration went on. The high-level tasks of EVis are to evaluate known and discover unknown triggers for events. To achieve these goals, the scientists need to: (1) rapidly explore relationships among time (both calendar and diurnal time), one or more E-conditions, and events (e.g., the **foundational analyses**), such as learning conditions coupled with high cracking rates and examining cracking rates at

different value ranges; (2) effectively explore temporally lagging relationships between E-conditions and events, such as identifying typical PE-conditions associated with high cracking rates and analyzing events whose PE-conditions were different from prior knowledge of the scientists; (3) interactively conduct intuition-based exploration of hypotheses (e.g., the **exploratory analyses**), such as highlighting data items with characteristics of interest and then examining their distributions in other dimensions and their relationships to events.

High-level requirements include the ability to: (1) employ vetted visualization practices of Earth scientists and leverage them; (2) effectively and efficiently examine—without coding—a large number of relationships, no matter if they are hypothesized ahead of visualizing the data or not; (3) examine intuitive visualizations through easy-to-use interactions since Earth scientists are usually not visualization experts. More minor specific requirements included the ability to: (1) visualize E-conditions with events, together with E-conditions without events, to eliminate biases, since a value range may contain a large/small number of events simply because the E-conditions are heavily/rarely sampled in this value range; (2) analyze events and environmental factors in the context of calendar time and diurnal time for seasonal and diurnal patterns; (3) provide quantitative information, such as event statistics, to guide visual exploration.

4 EVIS

4.1 Data Preprocessing and Concepts

To generate visualizations, the continuously monitored E-conditions were aggregated using a domain-expert-selected time granularity (one hour in this paper). This resulted in a time series dataset where the dimensions are environmental variables, and each data item recorded the aggregated values of these variables in an hour (thus a data item is called a **discrete-time-point**). Then, metrics for the natural phenomenon under investigation (e.g. sum of energy or count of rock cracking events in this paper) were aggregated using the same granularity and added to the dataset as a **Dimension of Interest (DOI)**. Discrete-time-points with no event detected are called **non-event hours** and those with non-zero event counts are called **event hours**. The New Mexico Rock dataset contains 26,304 discrete-time-points (the null-value missing data was linearly interpolated), including 2,240 event hours. Users could set either event energy or event count as the DOI in an exploration. The discrete-time-points users selected in exploration are called **Points Of Interest (POIs)**. The total number of event counts (or the energy sum) divided by the total number of hours of a group of discrete-time-points is called the **event rate** of the group.

4.2 Visualization Overview

EVis provides a set of coordinated visualizations (see Fig. 1 for a screenshot of EVis). To conduct foundational analyses that are common across virtually all Earth science applications, EVis provides scatterplots (Fig. 1(a)) to depict how individual environmental variables change over time and how they are related to the DOI, histograms (Fig. 1(c)) to show total numbers of event hours and event rates of discrete-time-points in different value ranges on each dimension, and heatmaps (Fig. 2(a)) to depict event rates of discrete-time-points in value ranges defined by three environmental variables. Although not yet widely employed in Earth sciences, EVis includes RadViz visualization [25,33] (Fig. 1(b)) because of its strong potential in this application [38], particularly when examining the relationships between the DOI and multiple variables and their temporal trends. The new approach **RadViz-Leash** is implemented in EVis for visually analyzing temporally lagging relationships among the DOI and multiple variables.

To coordinate and simplify visualizations for users, the same bubble metaphor and color coding are implemented in most visualizations. In particular, discrete-time-points are represented by bubbles in scatterplots, RadViz, and RadViz-Leash, whose sizes represent the DOI values of the discrete-time-points. The bigger the values, the bigger the bubbles. Event hours are represented by blue (unselected), orange (selected), or green (selected with a different approach) bubbles, and non-event hours are represented by small, hollow gray bubbles.

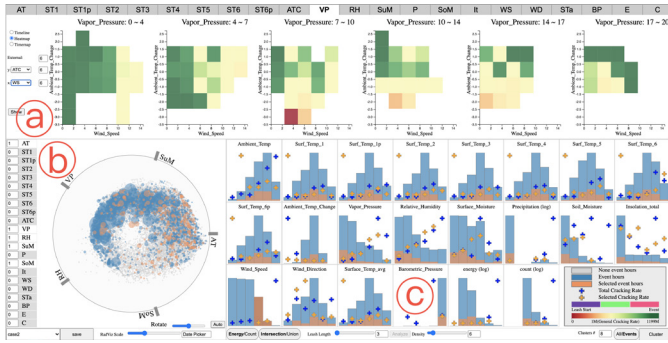


Fig. 2. Studying POIs with high wind speeds. (a) Heatmaps with Vapor Pressure (outer dimension), ATC (Y axis), and Wind Speed (X axis). (b) RadViz with a few well-known triggers as anchor dimensions. Highlighted POIs are well-distributed in this view. (c) Histograms. Bars of high wind speeds are selected.

EVIS supports intuition-based exploratory analyses and a smooth, flexible visual exploration pipeline. Users can select discrete-time-points from scatterplots, histograms, and RadViz. The selected discrete-time-points (POIs) are highlighted in all scatterplots and RadViz so users can examine their distributions in these views. Users can also examine aggregated information of the POIs in the histograms. Moreover, users can explore PE-conditions of the POIs within a user-selected time span at multiple levels of detail and examine discrete-time-points with similar PE-conditions of a focus POI using RadViz-Leash.

4.3 Basic Graphical Charts

Value-time scatterplot: Discrete-time-points are mapped to a 2D ‘value vs. time’ scatterplot as bubbles (Fig. 1(a)). The X and Y axes of the scatterplot are time and an environment variable, respectively. Users can observe the relationships among the DOI, the variable, and time via the positions, colors, and sizes of the bubbles. Users can click the tags above the scatterplot to switch among different environment variables. We chose this simple approach since it is intuitive but reveals rich insights. Line charts were not used since the many zero values in the dataset would lead to very cluttered line charts.

Date-diurnal time scatterplot: Discrete-time-points are also mapped to a 2D ‘date vs diurnal time’ scatterplot, inspired by LastHistory [7], to reveal diurnal patterns from midnight to midnight, which is of interest to scientists (Fig. 3(a)). The X and Y axes of this view are date and diurnal time, respectively.

Heatmaps: The heatmaps (Fig. 2(a)) provide insights into multivariate relationships among the DOI and three user-selected environmental variables. The variables are mapped to X, Y, and an outer dimension of the heatmaps, respectively. The outer dimension is divided into multiple value ranges, each for a heatmap. The X and Y dimensions are also divided into multiple value ranges. In each heatmap, the color of a cell represents the event rate of all discrete-time-points falling in the 3-dimensional region defined by the corresponding X, Y, and outer value range. The depicted color scale for event rate is normalized to the event rate of the whole dataset: green/red colors represent relatively low/high event rates. Users can interactively change the divisions of the value ranges to examine event rates at different granularity. The side-by-side heatmap layout allows users to compare event rates in different value ranges of the outer dimension easily.

Since E-conditions are not evenly distributed (e.g., there are very few hours with ambient temperatures lower than -10 Celsius degree), it is important for scientists to learn the densities of discrete-time-points in the cells. To support this task, EVIS overlays a scatterplot over the heatmap, whose X and Y axes are the same as the heatmap, and only discrete-time-points falling in the value range of the outer dimension for that heatmap are displayed as bubbles. Users can turn the scatterplots on or off by clicking a button. To reduce the complexity of the interface, the scatterplots and the heatmaps share the same display space. Users can switch among them by clicking a button.

Histograms: EVIS provides a set of histograms, one for each variable, to facilitate selection and reveal relationships between the DOI and the

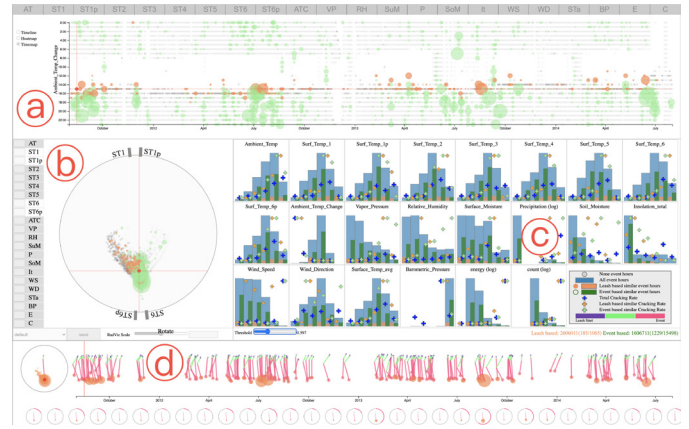


Fig. 3. Studying a leash of interest using similarity search. Red: the focus POI. Orange: POIs with similar leashes to the focus POI. Green: POIs with dissimilar leashes but similar POI positions to the focus POI. (a) The date-diurnal time scatterplot shows orange POIs are distributed around 15:00 pm. (b) RadViz shows different distributions of orange POIs and green POIs. (c) The histograms show orange POIs and green POIs have different event rates in many value ranges. (d) Leashes of orange POIs are displayed in a timeline view and a detail view similar to Fig. 7 (not fully displayed due to the page limit).

variables with relation to the whole dataset and POIs. A histogram consists of a set of bars and crosses (Fig. 1(c) and 2(c)). Its X-axis is divided into equal-width value ranges of the variable it represents. Heights/Y positions of blue bars/crosses represent numbers of event hours/event rates of the whole dataset in the corresponding value ranges, normalized within each histogram to make the best use of the space. Numbers of event hours/event rates of POIs in the value ranges are represented by orange bars/crosses so that users can compare them with those of the whole dataset. A bar/cross is hidden if the number of event hours/rate is zero to distinguish it from those with small values.

Selection: Users can drag a rectangle to select discrete-time-points in any of the scatterplots. They can also click a bar of a histogram to select discrete-time-points in that value range. Users can merge a new selection with a previous selection via an intersection or union set operation. For example, a user selects discrete-time-points with low ambient temperatures from the Ambient Temperature histogram, and then intersects it with a selection from a rectangle in the date-diurnal time scatterplot to highlight discrete-time-points with low ambient temperatures that also occurred around noon. The POIs are highlighted in other visualizations, so the user can learn how the POIs are related to Wind Speed, Precipitation, and other variables. They may explain why the event counts/energy are different within similar ambient temperatures and times of the day.

4.4 RadViz

EVIS provides a RadViz view (Fig. 1(b) and 2(b)) to explore relationships among the DOI and multiple user-selected environmental variables. These variables are interactively selected and placed along the circumference of a circle as anchor dimensions. Discrete-time-points are projected to the interior of this circle as bubbles, whose sizes and positions reveal relationships among the DOI and the anchor dimensions. Assume $\vec{s}_1, \dots, \vec{s}_n$ are vectors for anchor positions of n dimensions. After normalizing all dimensions to $[0, 1]$, a discrete-time-point $[d_1, \dots, d_n]$'s projection vector \vec{r} is calculated using the following equation:

$$\vec{r} = \left(\sum_{j=1}^n \vec{s}_j d_j \right) / \left(\sum_{j=1}^n d_j \right) \quad (1)$$

We chose RadViz over other dimension reduction methods because (1) RadViz explicitly reveals the relationships between the discrete-time-points and the anchor dimensions and (2) the projections are solely decided by the anchor positions and the domains of the anchor dimensions and, thus, we can use RadViz to generate trajectories of multivariate time series (see Section 4.5).

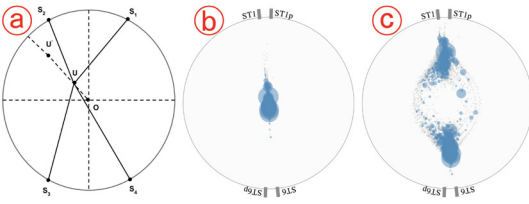


Fig. 4. Rescale the projections. (a) An illustration of how rescaling works. (b) A cluttered RadViz view. (c) (b) rescaled with $s = 5$.

A drawback of RadViz is that the projections are often concentrated in a small region of the 2D space. To overcome this problem, we propose two new interactive distortion methods for RadViz. Note that there are automated anchor dimension relocation algorithms [4, 12] for clutter reduction in RadViz. We did not choose them because they would change the anchor layout created by scientists, which may carry their domain knowledge and hypotheses. In addition, our distortion methods are fully controllable by users, which makes it easier for users to understand the distortion.

Distortion method 1: Rescale the projections If the anchor dimensions are highly correlated, the projections may be concentrated near the center of the circle, which makes RadViz less effective in revealing subtle patterns. For example, Fig. 4(b) shows a RadViz view where anchor dimensions are top and bottom surface temperatures. These temperatures are highly correlated and all the points are concentrated at the center. Such clutter cannot be reduced by reallocating anchor positions using algorithms such as RadViz Deluxe [12]. To address this problem, we introduce a global rescale factor s to move \vec{u} to \vec{u}' using the following equation:

$$\vec{u}' = \frac{\vec{u}}{\|\vec{u}\|} \cdot \|\vec{u}\|^{\frac{1}{s}}, (s \geq 1) \quad (2)$$

Since $0 \leq \|\vec{u}\| \leq 1$, increasing s will monotonically push \vec{u}' away from the center without changing its original direction. If $\|\vec{u}_1\| < \|\vec{u}_2\|$, then $\|\vec{u}_1'\| < \|\vec{u}_2'\|$ for all $s \geq 1$. The smaller $\|\vec{u}\|$, the bigger the distortion. Therefore, this distortion amplifies tiny distances among data points around the center of RadViz while keeping their angles and orders in the radial coordinate unchanged. Users can interactively change s through a scale to enlarge or reduce the distortion effect. Fig. 4(a) illustrates how u moves to u' with rescaling. Fig. 4(c) shows Fig. 4(b) rescaled with $s = 5$.

Distortion method 2: Adjust weights of dimensions It is often observed that data points are concentrated in a region distant or proximal to an anchor dimension in RadViz (see Fig. 5(b) for an example). We propose a new approach to pull data points toward or push them away from an anchor dimension to reduce clutter. Inspired by iPCA [26] where users can interactively set weights for dimensions in a PCA, we allow users to interactively assign a non-negative weight w_j to an anchor dimension j . The deformed projection \vec{u}' is calculated with the following equation:

$$\vec{u}' = \left(\sum_{j=1}^n \vec{s}_j w_j d_j \right) / \left(\sum_{j=1}^n w_j d_j \right) \quad (3)$$

Equation 1 is a special case of equation 3 where all weights are 1. Assume we adjust w_1 and keep the other weights 1. It is trivial to see that if $d_1 = 0$, \vec{u}' will be the same as \vec{u} , the position calculated with equation 1. If $d_1 > 0$ (all dimensions are normalized to $[0,1]$), by defining $\alpha = \frac{\sum_{j=2}^n d_j}{d_1}$, we get the displacement from \vec{u} to \vec{u}' with the following equation:

$$\vec{u}' - \vec{u} = \frac{w_1 - 1}{w_1 + \alpha} (\vec{s}_1 - \vec{u}) \quad (4)$$

It shows that when $w_1 > 1$ or $w_1 < 1$, all the projections will move toward or away from the anchor position of dimension 1 along straight lines connecting the anchor position and the original projections. It makes the distortion results visually easy to follow. The smaller α , the bigger the movement. Thus, increasing the weight will amplify the influence of a dimension in RadViz. When $w_1 = 0$, dimension 1 has no influence on RadViz according to equation 3.

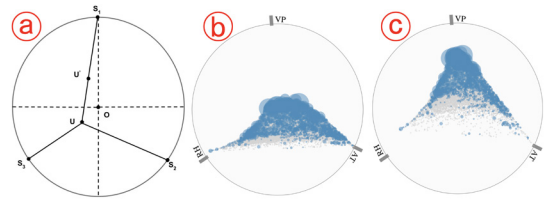


Fig. 5. Adjust the weight of a dimension. (a) An illustration of how the distortion works when the weight of the dimension on the top is increased. (b) A skewed RadViz view. (c) (b) after the weight of the dimension on the top is increased to $w = 5$.

Fig. 5(a) illustrates how u moves to u' after the weight of the dimension on the top is increased. Fig. 5(c) shows Fig. 5(b) after increasing the weight of the top dimension from 1 to 5.

Besides the distortion techniques, EViz allows users to **interactively add/remove dimensions** to/from RadViz and **manually relocate the dimension anchors** by dragging the anchors around the circle. The discrete-time-point projections will be changed accordingly. After a manual relocation, users can use the **auto-adjustment** function to evenly distribute the anchors around the circle without changing their order to get a balanced view. If users do not like the angles of the anchors, they can **rotate** the whole RadViz view by dragging a scale for preferred orientations.

By interactively selecting dimensions for RadViz and changing their anchor positions, scientists can experiment with different RadViz setups for hypothesis testing. For example, a scientist hypothesized that the larger the differences between the top and bottom surface temperatures of a rock, the more frequently the rock will crack. Thus, she placed the top/bottom surface temperatures to the top/bottom of the RadViz circle, as shown in Fig. 1(b). The distribution of large bubbles verified her hypothesis. Scientists can also experiment with different layouts and keep the layouts with interesting patterns, which may reveal novel relationships they have not noticed before.

Users can interactively **select** discrete-time-points from RadViz using a rectangle or a selection box with an arbitrary shape. Since dimension reduction always causes information loss, examining the POIs in other views helps users get a more precise picture of them.

4.5 RadViz-Leash

RadViz-Leash aims to allow scientists to visually analyze temporally lagging relationships among the DOI and multiple variables. It projects the time series ahead of a POI, which consists of the anchor dimensions of RadViz within a user-selected time span, as a trajectory on RadViz. Because the time series is temporally 'in front of' the POI, the trajectory is called a 'leash' of the POI (Fig. 6(a) and (b)). The geometric shape and position of the leash visually depict the changing PE-conditions of the POI within the time span. To analyze PE-conditions of a large number of POIs, RadViz-Leash automatically clusters the POIs based on their leash similarity. The clusters are visually presented to users in multiple views, where users can visually examine groupings of the most commonly recurring PE-conditions associated with the POIs, as well as their relationships to the DOI.

Projection and Leash Visualization: Denote a POI as $p_t = (d_{1,t}, d_{2,t}, \dots, d_{n,t})$, where n is the number of anchor dimensions and t is the time stamp of the POI. The PE-conditions (anchor dimensions only) of p_t in a time span of length k is a multivariate time series $\{p_{t-k}, \dots, p_{t-1}, p_t\}$. To visualize this time series, RadViz-Leash projects $p_{t-k}, \dots, p_{t-1}, p_t$ to the RadViz space. The projections are denoted as $p'_{t-k}, \dots, p'_{t-1}, p'_t$. RadViz-Leash connects $p'_{t-k}, \dots, p'_{t-1}, p'_t$ in their temporal order using line segments and draws a bubble at p'_t to represent the DOI at the POI. Colors of the line segments indicate their temporal distances to the POI. The resulting visualization is a leash representing the POI and its PE-conditions (see Fig. 6(a) and (b)). Fig. 6 shows a set of leashes with $k = 24$ hours. The lengths and directions of the line segments in a leash provide rich information about how the PE-conditions of a POI evolve over time. For example, the PE-conditions of the POIs in the third row changed much more significantly than those of the

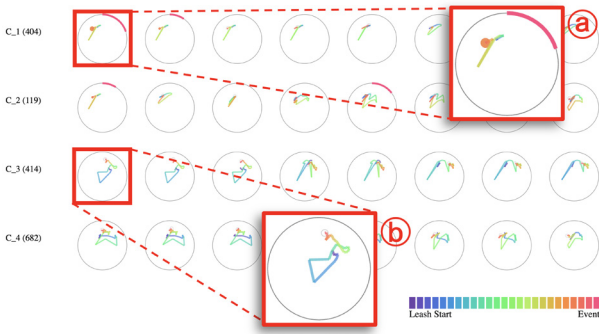


Fig. 6. A portion of the cluster sequential view (sorted by leash similarity) of POIs at freezing temperatures (leash length = 24 hours). (a) and (b) are zoomed-in views of leashes from different clusters. The POI in (a) is an event hour and the POI in (b) is a non-event hour.

POIs in the first row of Fig. 6.

Distance Calculation: RadViz-Leash clusters POIs based on leash distance rather than PE-condition distance in the raw data space. In this way, clustering results are consistent with what users see from the visualization, and domain knowledge integrated into the dimension reduction process is preserved. To calculate leash distance, we borrowed ideas from trajectory distance metrics. Magdy et al. [30] presented a survey on measures for trajectory distance between two moving objects. Measures considering both spatial information and speeds of the moving objects were unnecessarily complex for our application since our trajectories have uniformed time steps. Because the absolute positions of the leashes in a RadViz view are important, we excluded spatial assembling distance [10] and angular metric for shape distance [32] from consideration because they are not strict in the distance of absolute positions. Hausdorff and Frechet distance [3] did not have this drawback. It defined the distance of two trajectories A and B using the following equation:

$$H(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (5)$$

This distance captured how dissimilar two undirected trajectories are based on their geometry. However, it was not strict in the directions of the trajectories. For example, the distance of two leashes with the same geometry but with opposite directions, indicating conditions changing in opposite directions in our application, would be '0' with Equation 5. This was misleading in our application. Thus, we decided to modify Hausdorff and Frechet distance to take the temporal order of the leashes into consideration. Denoting leash A as $\{a_k, a_{k-1}, \dots, a_0\}$ and leash B as $\{b_k, b_{k-1}, \dots, b_0\}$, we calculate the distance between leash A and leash B $\delta(A, B)$ as follows:

$$\delta(A, B) = \max(\delta(A \rightarrow B), \delta(B \rightarrow A)) \quad (6)$$

$$\delta(A \rightarrow B) = \max_{a_i \in A} \min_{b_j \in B} \|a_i - b_j\| \quad (7)$$

$B[i - \phi, i + \phi]$ is a subsequence of B with the range of index from $i - \phi$ to $i + \phi$. $\phi = 0$, if $0 < i < 4$; $\phi = 1$, if $4 < i < 7$; $\phi = 2$, if $i > 7$. $\delta(A, B)$ is decided by the maximum of the point to point distances between the two leashes. In general, a point in a leash should be compared with the point with the same temporal order in the other leash. However, it is too strict for long leashes. Thus, equation 7 relaxes this requirement by ϕ to compare a point with points appearing a little sooner or later in the other leash and use the minimum distance when the point is temporally far away from the POI. The parameters used here are set by expert users based on domain knowledge.

Clustering: RadViz-Leash uses k-means clustering [24] to group a set of POIs into a user-defined number of clusters based on their leash distances. The centroid of a cluster in each iteration is defined as the POI with the minimum sum of distances to all other POIs in the same cluster. We chose k-means over other clustering algorithms since (1) it minimizes within-cluster variances, which is a desired feature for the POI clusters. Clustering algorithms generating non-spherically shaped clusters are not suitable in our application since they may group leashes representing quite different PE-conditions together; (2) it allows scientists to flexibly set the number of clusters to examine; and (3) it is fast enough for interactive visual exploration.

Authorized licensed use limited to: University of North Carolina at Charlotte. Downloaded on December 19, 2025 at 20:45:42 UTC from IEEE Xplore. Restrictions apply.

Overview: Overviews of clustering results are provided to allow users to: (1) browse typical PE-conditions associated with POIs and learn their relationships with the DOI, anchor dimensions, and time; (2) identify PE-conditions of interest for further examination. We chose to display PE-conditions in the overviews as leashes since they are compact and information-rich. Two alternative designs were considered: (1) to display all or a cluster of POIs and their leashes in the same RadViz view; (2) to display each POI and its leash in their own RadViz view using small graphics. We chose Design 2 since Design 1 led to a heavily cluttered RadViz view. In addition, the layout of the small graphics can reveal useful information to users.

Multiple layout strategies were tested and two were chosen in the final prototype. One is the **cluster timeline view**. It emphasizes relationships between the clusters and time. It places the small graphics on a set of parallel, horizontal timelines, each for a cluster (see Fig. 1(d)). The circles of RadViz are hidden, but the centers of the circles of the same cluster are vertically aligned to allow users to compare the vertical positions of the leashes in RadViz. The X positions of the endpoints of the leashes indicate the timestamps of the POIs. This view allows users to examine and compare leash colors, leash shapes, and time frequencies of different clusters in a compact display. Since it is important to compare event rates of different clusters, interactions are provided to allow users to **sort** the clusters by event rate, **dim** leashes of non-event hours to highlight leashes of event hours, and read event rates which are displayed as text on the left end of the timelines (see Fig. 1(d)).

In our initial design, we displayed the circles to provide a better picture of relative positions of leashes in RadViz. Since the circles caused severe clutter and distracted users from observing the colors and shapes of the leashes, they were removed from the final design. We experimented using the Y positions of the circle centers to reveal the similarity between a POI and the center of its cluster or a user-selected POI. This layout was confusing after the circles were removed. We also tried mapping the timestamps of the POIs to the X positions of the circle centers, but since the circle centers are not displayed, users could not judge when an event happened with this design.

Although the cluster timeline view is space-efficient and presents a nice overview of the clusters, it is difficult to select and compare individual leashes in it. To address this problem, a **zoom** function is provided. It allows users to drag and drop a rectangle on this view. All leashes in the rectangle will be displayed, together with their circles, in a pop-up window without overlaps. Users can create multiple pop-up windows at the same time to compare leashes in different time ranges or different clusters. However, it is trivial to do so. Thus, the second layout, the **cluster sequential view** is provided in the final prototype.

The **cluster sequential view** keeps the row-by-row layout of clusters in the timeline view so that users can switch between the two views without losing the positions of the clusters. In each row, the leashes and their circles are displayed sequentially without overlaps. Users need to horizontally scroll the screen if there are more POIs in a cluster than the screen width can hold. To reduce the need for scrolling and highlight important features of a cluster, sorting and filtering interactions are provided. Users can **sort** the leashes by **sum of DOIs of all discrete-time-points in the leashes** in descending order since scientists are interested in PE-conditions of significant events, or by **leash similarity**, so that similar leashes are placed adjacent to each other to help users observe variations within a cluster (the sorting algorithm presented in [46] is used to minimize the total distance between adjacent leashes). A cluster may contain many POIs with redundant leashes, namely those leashes only shifting one or two hours in time. Inspired by Schreck et al. [39], a **filtering** interaction is provided to filter out POIs with redundant leashes using a user-defined time subinterval. Only the latest POI among consecutive POIs within the time subinterval is displayed in a cluster.

There are other possible ways to lay out the leashes to form an overview (see [44] for a survey of glyph placement). For example, all leashes can be mapped to a 2D space where the distances between two leashes reveal the similarity between them. However, they may introduce extra complexities and clutter to the visualization.

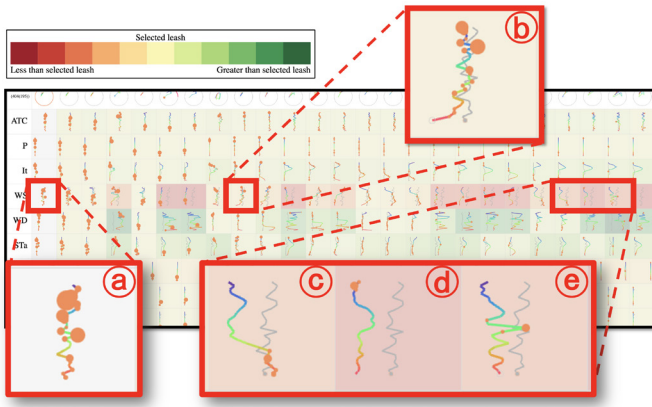


Fig. 7. A portion of the detail view of a cluster. The first column is selected for comparison. (a)-(e) are zoomed-in views of several line charts of the Wind Speed variable. (a) is the chart of the selected POI. (b)-(e) compare (a) (gray lines) with charts of other POIs. They reveal high wind speeds might be related to events.

Detail View: To understand relationships between events and PE-conditions, scientists need to examine and compare POIs and their PE-conditions in full detail. To support this task, a detail view is provided in the prototype. It is a matrix of small graphics to show leashes and detailed PE-conditions of a group of POIs (see Figure 7). In the matrix, each column displays a POI and its PE-conditions. The columns can be sorted and filtered by the methods introduced in the cluster sequential view. The first row displays the leashes of the POIs. Each of the other rows displays line charts [2] of the POIs on an environmental variable in small graphics. The anchor dimensions of RadViz are displayed on the top, followed by non-anchor dimensions. In the line charts, the Y axes are time axes, with the time of the starting points of the leashes at the top and the POIs at the bottom. The Y values of the line are double encoded using colors of the lines, in the same way as the leash segments. The X-axes represent the variables with the smallest values on the left. Note the X-axes in the same row have the same scale for all columns while the Y axes are aligned by the timestamps of the POIs.

Since each line chart only shows one time series on one dimension, we can map the DOI values of all event hours in the leashes to bubbles without cluttering the line charts (see Figure 7(a)). Our domain experts commented that this is extremely useful, since the bubbles enable them to directly relate the evolving PE-conditions to all events that occurred during the periods of the leashes, which is of interest.

The detail view is useful for finding new event triggers by comparing POIs with similar leashes (selected by clustering or similarity search). Those POIs have similar temporal patterns on the anchor dimensions, which may be selected by scientists since they are known triggers. Why would a POI/leash have different DOI values from other POIs/leashes? The reason might be that they behave differently on non-anchor dimensions, which may be triggers not noticed by the scientists. The detail view provides a **comparison** interaction to facilitate this type of analyses. In particular, users can set a POI as a focus by clicking its leash, and compare it with other POIs on all the dimensions (see Fig. 7). In the comparison mode, the line charts in focus are copied to other line charts in the same rows and displayed in gray for a direct comparison (see Fig. 7(b)-(e)). The background color of the other line charts indicates the difference. Red means the focus has higher average values and green means the focus has lower average values on that dimension. The larger the difference, the darker the color. The background colors allow users to capture insights from a large matrix at a glance. In Fig. 7, the dark red background reveals a focus with large events having higher wind speeds than most POIs with small or no events, which inspires the scientists that wind speed might be a trigger for rock cracking.

An alternative design to the line chart matrix was to overlay all the line charts of the same dimensions in the same view. This design was

not used since the view was cluttered and it was difficult for users to examine and compare multiple leashes in detail with the overlapping line charts. Another alternative design was to map the time series to timelines, one for each POI on each dimension, whose colors indicated the values of the PE-conditions. This design was not used because color was used in leashes to encode the time dimension. Using color in a different way confused users.

The detail view can be triggered by clicking a cluster name in the overviews to examine POIs in the cluster. It can also be triggered by clicking a POI to set it as a focus, and then setting a leash distance threshold d to select POIs with similar leashes to the focus (leash distance $< d$) from the whole dataset. In this **similarity search mode**, a new window is opened (see Fig. 3), which contains basic charts (Fig. 3(a) and (c)), RadViz (Fig. 3(b)), a timeline view (Fig. 3(d)) and a detail view for the selected POIs. The basic charts and RadViz allow users to compare selected POIs against POIs with similar conditions on the anchor dimensions when the events happened (POI projection distance $< d$), but dissimilar PE-conditions on the anchor dimensions (leash distance $\geq d$). They are displayed in green (event hours) or light gray (non-event hours) to provide context to selected POIs (orange for event hours and dark gray for non-event hours). Other discrete-time-points are hidden in this window to reduce clutter.

5 CASE STUDIES

We illustrate the usefulness of EVis using two case studies. They were conducted on the New Mexico Rock dataset [17] by the senior scientist and two vis-researchers from our team via several Zoom meetings (all recorded). In the meetings, a vis-researcher ran EVis on his PC and shared his screen. The scientist orally instructed the vis-researcher what she wanted to do and why, while drawing on the screen with the Zoom annotation function to point to the places where she wanted to select or zoom in. She also verbalized what she found and which hypotheses she had. The other vis-researcher took notes during the meetings and summarized the case studies after the meetings based on the notes and the recordings. EVis settings used in the explorations were saved using a save function provided by EVis, so the results could be duplicated when writing this paper.

5.1 Case 1: Temperature and Rock Cracking

The scientist had discovered through past work that temperature is related to cracking [17]. However, there was a lack of detailed analysis—or full understanding—of this complicated relationship. Thus, she decided to use EVis to get more insights into how temperature is related to cracking events. Based on theoretical modeling [36], she hypothesized that rocks would crack when their top and bottom temperatures are diverging in their rate of temperature change. Previously, she had only employed surface air temperature change as a proxy for this effect but had not observed it directly in the dataset. So, using EVis, she manually set up four anchor dimensions to the RadViz view. Surface_Temperature_1 and Surface_Temperature_1p, which record top surface temperatures of the rock, were placed on the top of RadViz. Surface_Temperature_6 and Surface_Temperature_6p, which record bottom surface temperatures of the rock, were placed on the bottom of RadViz (Fig. 1(b)). The discrete-time-points were concentrated around the center. Thus, she distorted the projections to separate the points cluttered at the center. From Fig. 1(b), she observed that most hours with large DOIs are either high or low in the circle, which meant that the top surface temperatures were much higher or much lower than the bottom temperatures. This supported her discovery from the previous study, as well as the theoretical models.

According to her previous research, high cracking rates correlate with negative Ambient Temperature Change (ATC). A negative ATC value means a large per-minute drop in temperature within the discrete-time-point. To explore more details, she opened the value-time scatterplot for ATC (Fig. 1(a)). She saw large bubbles (hours with high DOI values) with negative ATC values. She dragged a rectangle in the scatterplot to select discrete-time-points in that region (Fig. 1(a)). She scanned the histograms (Fig. 1(c)) for patterns of the selected hours. According to the orange crosses in the histogram, the POIs have

much higher event rates than the whole dataset in most value ranges of variables such as Vapor Pressure and Temperature. It confirmed her prior knowledge that multiple variables align during high DOI periods.

Next, she explored the temporally-lagging relationships between the top and bottom surface temperatures of the POIs. She set a leash length of three hours with a cluster number of eight, based on her prior knowledge of expected lag-importance and the data selected. Then, she clustered the POIs. From the cluster timeline view (Fig. 1(d)), she examined the leashes of each cluster and checked their event rates from both the text and the ratio between bright leashes (event hours) and dimmed leashes (non-event hours). She found several clusters with long leashes, indicating the temperatures changed significantly within the three hours before the selected hours. Interestingly, she found that after the temperatures changed significantly from a situation where top temperatures were lower than bottom temperatures to a situation where top temperatures were higher than bottom temperatures (C_7), the cracking rates in both energy and counts were much lower than when they changed in the opposite direction (C_0, C_1, C_2). She also noticed that these long leash clusters are dominated by different colors, which indicate different temporal patterns. For example, most POIs in C_2 were preceded by relatively slow changes in the first two hours (short purple and light green lines) and then rapid changes in the third hour (long red lines). Identifying these clusters of different, specific, lagging conditions leading to events was a novel and impactful finding from the dataset.

To further explore one of the clusters, the scientist set a leash from C_2 as a sample and opened the similarity search view to further analyze it (Fig. 3). In this view, she set a similarity threshold. All event/non-event hours whose leash distances to the sample are less than the threshold are displayed as orange/dark gray circles. Other event/non-event hours whose distance to the event hour of the sample leash in the RadViz view are displayed in green/light gray circles. From the date-diurnal time scatterplot (Fig. 3(a)), it can be seen that the orange circles, namely the event hours with leashes of the pattern of interest, are distributed close to the daily time of 15:00 pm. The green circles do not have this pattern. This result supported the idea that leashes can better describe the characteristics of cracking events than discrete-time-points. In addition, from the histograms (Fig. 3(c)), she found that the orange set has different event rates from the green set in multiple bars. The discrete-time-points, however, have similar conditions themselves in the anchor dimensions – again a novel finding.

5.2 Case 2: Exploratory Analyses and Unexpected Patterns about Wind Speed

The scientist conducted another exploratory analysis on the New Mexico Rock dataset [17]. Her initial interest was in events with low to sub-freezing ambient temperatures, a condition well-documented in her field to produce cracking. Thus, she selected those discrete-time-points from the Ambient Temperature histogram and set up RadViz with Vapor Pressure, Relative Humidity, Ambient Temperature, Surface Moisture, and Soil Moisture which were hypothesized to influence cracking rates. Then, she set the leash length to 24 hours and conducted clustering to examine how event rates are related to leash shapes in this RadViz space. The clustering results are shown in Fig. 6. Cluster C_1 with small leashes but big events triggered her attention. She opened its detail view (Fig. 7). Sorting the leashes by total event energy in 24 hours, she noticed multiple leashes with many large bubbles on them in the first several columns. As expected, their conditions with respect to the anchor dimensions in 24 hours are fairly stable, since their leashes are small in RadViz. This is similar to other members of this cluster with small or no events. Are there any other triggers for those big events? She clicked a leash with many large bubbles (Fig 7(a)) to compare its line charts with other columns (Fig 7(b)-(e)). A row with light backgrounds for event-heavy time series and dark backgrounds for few/no event time series was immediately visible. It was the Wind Speed dimension. She zoomed in to this row, and found that high wind speeds are almost always coupled with high cracking rates under the selected low ambient temperatures.

The scientist knew that negative temperature changes are related to cracking. She suspected those high cracking rates were related to temperature drops caused by strong winds. She opened the heatmaps to test her hypothesis. She set Wind Speed as the X-axis of the heatmaps, Ambient Temperature Change as the Y-axis, and Vapor Pressure, another strong event trigger, as the outer dimension (Fig. 2(a)). The heatmaps indicate that with low vapor pressures, wind speed has a clear correlation with event rate. Then, she opened the scatterplots overlaying on the heatmaps. To her surprise, she found POIs with big events under conditions of high wind speed and positive ambient temperature changes from the heatmap of the lowest vapor pressures. This was different from her understanding. Now she learned that temperatures may increase when the winds are strong. Thus, the cracking that occurred in strong winds may also be related to some thermal mechanisms.

Were the selected high-wind event hours caused by antecedent temporal changes of the triggers she knew? She added ambient temperature change into the RadViz view and created another selection of discrete-time-points with high wind speeds. She observed that event hours are well distributed in the RadViz view (Fig. 2(b)). In other words, there are event hours with high wind speeds no matter their values of the anchor dimensions are high or low. She ran the clustering algorithm and found all clusters had an extremely high event rate, no matter whether their leashes were long or short. The conclusion that can be drawn from this visualization is that cracking is not strongly related to antecedent temporal changes of the anchor dimensions. Finally, she opened the value-time scatterplot of wind speed. She found multiple non-event hours with high wind speed. This indicated that another hypothesis she had, namely that cracking could be caused by sands blown against the rock by strong winds, might also be wrong since otherwise, all high wind speed hours should have had events. The findings were unexpected and exciting to her. She decided to further investigate these relationships between high wind speeds and rock cracking.

Visualization Take-away for the Case Studies: The process of conducting the case studies was very exciting to the scientist. In a few hours, she was able to confirm multiple insights for which she and her collaborators and graduate students spent multiple years to discover, as well as to discover new insights she had not expected before the case studies. According to her, the visualizations and interactions are intuitive, and the combination of them provide novel and powerful functions never feasible to her in the past. Though the interface has multiple views, she readily and continuously employed needed information from multiple portions of the interface—providing evidence that the interface density is warranted. She commented that both distortion methods are intuitive and useful. She liked that the second distortion method allowed her to strengthen the impact of critical E-conditions, which was exactly what she wanted. It was observed that the coordinated visualizations enabled her to analyze her hypotheses following multiple exploration paths smoothly—she started her explorations from multiple views, depending on the hypotheses she had.

6 EXPERT FEEDBACK

We demonstrated EVIS to Earth scientists from a range of disciplines to learn whether EVIS was applicable and useful for their research, as well as to collect feedback for improving EVIS (their concern about existing visualizations has been addressed in this paper; new functions they requested will be the future work). Thirteen scientists attended a one-hour Zoom meeting with the co-authors. They were known to us to be actively collecting or working on complex time-series datasets. They represented sub-disciplines of climatology, geology, rock mechanics, physical geography, and ecology. They included a range of professions and career stages: late-stage Ph.D. students (2), post-doctoral researchers and early-career academics (3), mid-career academics (3), international governmental employees (1), full professors (2), and experienced industry professionals (2). In the Zoom meeting, the Earth scientist coauthor first introduced her rock mechanical weathering research for five minutes, and then gave a live demo of EVIS with the New Mexico Rock dataset [17] for 40 minutes. A 15-minute discussion about EVIS followed the demonstration. A Google Form questionnaire was distributed to the scientists before the demo started. Eleven out of the thirteen scientists returned their answers to us within four days.

after the meeting—there had been no requirement to submit the form to attend the demonstration. Below are the questions and answers of the questionnaire (the original words of the experts are quoted):

Is the software applicable in your field? Nine scientists answered yes to this question. They commented that EVis is “extremely relevant and applicable” to their fields of research. “Its unique capability to visualize high-density multi-variable time-series data is exciting and is directly applicable” to their projects dealing with processes such as landscape evolution and subsurface fracturing. It is “useful for testing and exploring hypothesized mechanistic linkages between an observed process in the field and variability in the environmental conditions that drive it”, or even for “revealing linkages that were not hypothesized ahead of the experiment”. One scientist wrote: “This software could totally be useful in my field (structural geology). Research projects in my field typically collected 20+ variables which are almost never visualized or compared due to not having software capable of doing this. Being able to show all variables across multi-year projects might show relationships that were not considered before.” Two scientists answered maybe since they either do not work with complex time series or do not conduct open-end exploratory analyses.

Do you think the visualizations and functions are useful in data exploration? Ten scientists answered yes to this question. They commented that “based on the demo EVis is a really novel and useful set of visualizations” and “is great for opening up the possibilities of seeing connections that one might not have been previously looking for”. They confirmed the usefulness of RadViz-Leash: “Rock properties are inherently linked but it’s very difficult to see how many trends co-vary. The plots of data leashes show how the complex strain fields vary leading up to deformation events, which is very difficult to visualize, even with models and lab analysis. This is often done as 2D models which lose a lot of the important details. The tails allow you to see what is happening and better interpret the data.” “RadViz is a useful quantitative visualization tool to explore relations within multi-variable data and in the time domain. The ‘leash’ function is extremely useful to help zoom in on specific data streams and their relationship with independent measures of process.” They also commented that the basic visualization charts and the interactions are beneficial. One scientist answered not sure since he doesn’t use complex time series.

What are the potential benefits/advantages (if any) of exploring data using this software compared with existing approaches researchers are using in your field? The scientists commented their work was done piecemeal and they were not aware of any visualization tools tailored to their needs. They wrote “EVis seems better customized to flexible numbers of variables and to providing a straightforward workflow compared to existing more generic software” and “while EVis does not prove causation, it can shed light on potential connections. Traditional tools require you to actively look for specific connections (or lack thereof) that you had already suspected.” They commented “EVis has a strong potential to attract many more of us” and “EVis has the potential to be a game changer”.

What are the limitations? Please provide your top 2 suggestions to improve the software The scientists provided a set of suggestions to improve the usability and functionality of EVis. Two commented that the detail view was busy with too many plots. To address this problem, we (1) added the dissimilarity background, (2) added the sorting interactions so that users can focus on more interesting POIs, and (3) added the filtering interaction to hide redundant leashes. The scientists also suggested extending the applicability of EVis from the following aspects: (1) add 3D spatial visualization; (2) add statistics and signal analysis capabilities, such as basic 1D spectral analysis (e.g., FFT and wavelets); (3) provide thorough documentation, tutorial videos, and a user manual with examples; (4) allow users to upload datasets with different time scales, provide build-in tools to extrapolate and interpolate data between points, and allow users to create new variables via calculations; and (5) provide output functions for further analysis and make EVis compatible with other software such as R. We will add these functions into EVis in the future.

Authorized licensed use limited to: University of North Carolina at Charlotte. Downloaded on December 19, 2025 at 20:45:42 UTC from IEEE Xplore. Restrictions apply.

7 DISCUSSIONS

Compared with popular dimension reduction methods, RadViz is less familiar to Earth scientists and, therefore, has a steeper learning curve. However, it supports required tasks better than existing commonly employed techniques. The projection of a data point generated by MDS [28, 29], PCA [20, 45], t-SNE [35], and UMAP [31] is influenced by other data points in the input. Therefore, leashes generated by them will change when the set of POIs changes. RadViz does not have this problem—the projection of a data point is independent of other data points. Moreover, RadViz allows users to examine data points/leashes in relation to their attributes, which is a desired feature for Earth scientists. PCA [45] and related displays, such as biplots [22], are not optimized for this task [12]. MDS [28, 29] and t-SNE [35] cannot support this task explicitly since they lose the data-dimension relationships in the projection process [12].

In EVis, users can interactively adjust RadViz, which will be propagated to leashes, using distortion and anchor relocation interactions. We believe these interactions are necessary and beneficial. First, without distortions, with an anchor setting as the one used in the first case study, most points are clustered at the center of RadViz, which makes insight discovery and leash clustering extremely difficult. Second, since distortions magnify subtle differences in data, RadViz-Leash can group data with those differences captured, which helps users capture insights hidden from other analysis approaches. Third, the process of interactively adjusting RadViz is a visual exploration process for insights—when users see interesting patterns in RadViz (e.g., big bubbles are clustered in RadViz), they can fix the layout and make selections from RadViz (e.g., selecting the big bubbles clustered) for further analyses from other views. This is a unique approach unavailable in existing practices. A limitation of our current approach is that users may need to experiment with multiple layouts to find interesting ones. In the future, we will study how to leverage this with automated approaches so that users can conduct explorations more efficiently.

8 CONCLUSION AND FUTURE WORK

In this paper, we propose EVis, a new visual analytics prototype for Earth scientists. It provides a rich set of visualizations and interactions for exploratory analyses of natural phenomena and their driving E-conditions. Our case studies with a domain expert illustrated the usefulness of EVis in rock mechanical weathering studies. Feedback from eleven scientists from varying sub-disciplines of Earth science revealed the broad applicability and potential usefulness of EVis in Earth science. Meanwhile, our practice of RadViz on Earth science further illustrates the functionality and uniqueness of this long-existing dimension reduction technique. Another major contribution of this paper is RadViz-Leash, a RadViz projection and clustering-based visual analytics approach for multivariate time series analyses. It can still be improved to further increase its scalability, but its usefulness has already been demonstrated in our case studies.

We believe EVis has the potential to bridge the gap between Earth scientists and their visualization needs. To be fully employed by scientists, input and output functions, insight management functions, signal analysis and statistics functions commonly used by Earth scientists, and more functions the domain experts suggested in Section 6 still need to be added into EVis. After adding those functions, we will evaluate EVis with datasets from different sub-disciplines of Earth science. In addition, we plan to develop prototypes for other disciplines, such as education, where temporally lagging relationships need to be analyzed using the techniques reported in this paper.

ACKNOWLEDGMENTS

We would like to thank the domain experts of sub-disciplines for their expert feedback. This publication is based upon work supported by the National Science Foundation under Grant Nos. EAR #00844335, #844401, and #0805277, and the University of North Carolina at Charlotte.

REFERENCES

[1] Tableau. <https://www.tableau.com/>.

- [2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of time-oriented data*. Springer Science & Business Media, 2011.
- [3] H. Alt. The computational geometry of comparing shapes. In S. Albers, H. Alt, and S. Näher, eds., *Efficient Algorithms: Essays Dedicated to Kurt Mehlhorn on the Occasion of His 60th Birthday*, pp. 235–248. Springer, 2009.
- [4] M. Angelini, G. Blasilli, S. Lenti, A. Palleschi, and G. Santucci. Towards enhancing radviz analysis and interpretation. In *2019 IEEE Visualization Conference (VIS)*, pp. 226–230, 2019.
- [5] R. Arias-Hernandez, L. T. Kaastra, T. M. Green, and B. Fisher. Pair analytics: Capturing reasoning processes in collaborative visual analytics. In *2011 44th Hawaii international conference on system sciences*, pp. 1–10, 2011.
- [6] B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, and P. Dragicevic. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):559–568, 2016.
- [7] D. Baur, F. Seiffert, M. Sedlmair, and S. Boring. The streams of our lives: Visualizing listening histories in context. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1119–1128, 2010.
- [8] M. Bostock, V. Ogievetsky, and J. Heer. D³ data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.
- [9] P. Buono, A. Aris, C. Plaisant, A. Khella, and B. Shneiderman. Interactive pattern search in time series. In *Visualization and Data Analysis 2005*, vol. 5669, pp. 175–186. International Society for Optics and Photonics, 2005.
- [10] Y. Chen, M. A. Nascimento, B. C. Ooi, and A. K. Tung. Spade: On shape-based pattern detection in streaming time series. In *2007 IEEE 23rd International Conference on Data Engineering*, pp. 786–795, 2007.
- [11] S. Cheng, K. Mueller, and W. Xu. A framework to visualize temporal behavioral relationships in streaming multivariate data. In *2016 New York Scientific Data Summit (NYSDS)*, pp. 1–10, 2016.
- [12] S. Cheng, W. Xu, and K. Mueller. Radviz deluxe: An attribute-aware display for multivariate data. *Processes*, 5(4):75, 2017.
- [13] T. Dasu, D. F. Swayne, and D. Poole. Grouping multivariate time series: A case study. In *Proceedings of the IEEE Workshop on Temporal Data Mining: Algorithms, Theory and Applications, in conjunction with the Conference on Data Mining, Houston*, pp. 25–32, 2005.
- [14] P. D’Urso and E. A. Maharaj. Wavelets-based clustering of multivariate time series. *Fuzzy Sets and Systems*, 193:33–61, 2012.
- [15] M. C. Eppes and R. Keanini. Mechanical weathering and rock erosion by climate-dependent subcritical cracking. *Reviews of Geophysics*, 55(2):470–508, 2017.
- [16] M. C. Eppes, B. Magi, B. Hallet, E. Delmelle, P. Mackenzie-Helnwein, K. Warren, and S. Swami. Deciphering the role of solar-induced thermal stresses in rock weathering. *Bulletin*, 128(9-10):1315–1338, 2016.
- [17] M. C. Eppes, B. Magi, J. Scheff, K. Warren, S. Ching, and T. Feng. Warmer, wetter climates accelerate mechanical weathering in field data, independent of stress-loading. *Geophysical Research Letters*, p. e2020GL089062, 2020.
- [18] M. C. Eppes, K. Warren, E. Hinson, and L. Dash. Long term monitoring of rock surface temperature and rock cracking in temperate and desert climates. *AGUFM*, 2012:EP41F-0848, 2012.
- [19] T.-c. Fu. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1):164–181, 2011.
- [20] T. Fujiwara, J. K. Chou, S. Shilpika, P. Xu, L. Ren, and K. L. Ma. An incremental dimensionality reduction method for visualizing streaming multidimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):418–428, 2020.
- [21] T. Fujiwara, N. Sakamoto, J. Nonaka, K. Yamamoto, K.-L. Ma, et al. A visual analytics framework for reviewing multivariate time-series data with dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1601–1611, 2020.
- [22] K. R. Gabriel. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3):453–467, 1971.
- [23] S. Ghassempour, F. Girosi, and A. Maeder. Clustering multivariate time series using hidden Markov models. *International journal of environmental research and public health*, 11(3):2741–2763, 2014.
- [24] J. A. Hartigan. *Clustering Algorithms*. John Wiley Sons, Inc., USA, 99th ed., 1975.
- [25] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley. DNA visual and analytic data mining. In *Proceedings. Visualization’97*, pp. 437–441, 1997.
- [26] D. H. Jeong, C. Ziemkiewicz, B. Fisher, W. Ribarsky, and R. Chang. iPCA: An interactive system for pca-based visual analytics. *Computer Graphics Forum*, 28(3):767–774, 2009.
- [27] T. Kohonen. *Self-organizing maps*. Springer Science & Business Media, 2012.
- [28] J. B. Kruskal and M. Wish. *Multidimensional scaling*. SAGE Publications, Inc., 1978.
- [29] J. H. Lee, K. T. McDonnell, A. Zelenyuk, D. Imre, and K. Mueller. A structure-based distance metric for high-dimensional space exploration with multidimensional scaling. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):351–364, 2014.
- [30] N. Magdy, M. A. Sakr, T. Mostafa, and K. El-Bahnasy. Review on trajectory similarity measures. In *2015 IEEE seventh international conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 613–619, 2015.
- [31] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- [32] T. Nakamura, K. Taki, H. Nomiya, K. Seki, and K. Uehara. A shape-based similarity measure for time series data with ensemble learning. *Pattern Analysis and Applications*, 16(4):535–548, 2013.
- [33] L. Novakova and O. Štepanková. RadViz and identification of clusters in multidimensional data. In *2009 13th International Conference Information Visualisation*, pp. 104–109, 2009.
- [34] Z. Orémuš, K. A. Hassan, J. Chmelík, M. Kňázková, J. Byška, R. G. Raidou, and B. Kozlíková. PINGU principles of interactive navigation for geospatial understanding. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 216–225, 2020.
- [35] P. E. Rauber, A. X. Falcão, and A. C. Telea. Visualizing time-dependent data using dynamic t-SNE. In *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization: Short Papers*, p. 73–77, 2016.
- [36] B. Ravaji, V. Ali-Lagoa, M. Delbo, and J. W. Wilkerson. Unraveling the mechanics of thermal stress weathering: Rate-effects, size-effects, and scaling laws. *Journal of Geophysical Research: Planets*, 124(12):3304–3328, 2019.
- [37] H. Roy, M. Pocock, C. Preston, D. Roy, J. Savage, J. Tweddle, and L. Robinson. Understanding citizen science & environmental monitoring. Final report on behalf of UK-EOF. NERC Centre for Ecology & Hydrology and Natural History Museum. *Natural History Museum, London, UK*. See <http://www.ukEOF.org.uk/co.citizen.aspx> (accessed 28/11/2012), 2012.
- [38] P. Sajadi, A. Singh, Y.-F. Sang, S. Mukherjee, and K. Chapi. Assessing the key drivers of stream network configuration dynamics for tectonically active drainage basins using multitemporal satellite imagery and statistical analyses. *Geocarto International*, pp. 1–32, 2021.
- [39] T. Schreck, T. Tekušová, J. Kohlhammer, and D. Fellner. Trajectory-based visual analysis of large financial time series data. *ACM SIGKDD Explorations Newsletter*, 9(2):30–37, 2007.
- [40] A. Singhal and D. E. Seborg. Clustering multivariate time-series data. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 19(8):427–438, 2005.
- [41] R. Takami and Y. Takama. Visual analytics interface for time series data based on trajectory manipulation. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 342–347, 2018.
- [42] S. Tilkov and S. Vinoski. Node.js: Using javascript to build high-performance network programs. *IEEE Internet Computing*, 14(6):80–83, 2010.
- [43] S. van den Elzen, D. Holten, J. Blaas, and J. J. van Wijk. Reducing snapshots to points: A visual analytics approach to dynamic network exploration. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):1–10, 2016.
- [44] M. O. Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3-4):194–210, 2002.
- [45] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.
- [46] J. Yang, W. Peng, M. O. Ward, and E. A. Rundensteiner. Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. In *IEEE Symposium on Information Visualization 2003*, pp. 105–112, 2003.
- [47] F. Zhou, W. Huang, Y. Zhao, Y. Shi, X. Liang, and X. Fan. ENTVis: A visual analytic tool for entropy-based network traffic anomaly detection. *IEEE Computer Graphics and Applications*, 35(6):42–50, 2015.