COPE: Interactive Exploration of Co-Occurrence Patterns in Spatial Time Series

Jie Li, Siming Chen, Kang Zhang, Gennady Andrienko, and Natalia Andrienko

Abstract—Spatial time series is a common type of data dealt with in many domains, such as economic statistics and environmental science. There have been many studies focusing on finding and analyzing various kinds of events in time series; the term 'event' refers to significant changes or occurrences of particular patterns formed by consecutive attribute values. We focus on a further step in event analysis: discover temporal relationship patterns between event locations, i.e., repeated cases when there is a specific temporal relationship (same time, before, or after) between events occurring at two locations. This can provide important clues for understanding the formation and spreading mechanisms of events and interdependencies among spatial locations. We propose a visual exploration framework COPE (Co-Occurrence Pattern Exploration), which allows users to extract events of interest from data and detect various co-occurrence patterns among them. Case studies and expert reviews were conducted to verify the effectiveness and scalability of COPE using two real-world datasets.

Index Terms—Co-occurrence patterns, spatiotemporal visualization, spatial time series, visual analytics

1 Introduction

Spatial time series represent temporal variation of attribute values at multiple spatial locations [5]. For a given location, the data specify the values attained by one or more attributes at different time moments or intervals, further jointly called 'time steps'. Spatial time series describe spatiotemporal evolution processes, which may extend over large areas and/or develop over long periods of time. This makes spatial time series an important data type in many fields, such as economics [45], environment studies [36], urban traffic [53], and many others.

Analyzing spatial time series has become an important research topic in visual analytics [8], [30], [36], [43]. Most existing studies work on visualizing spatiotemporal distribution, analyzing temporal trends, and detecting anomalies. As a part of time series analysis, some of the proposed approaches support detection and exploration of various spatiotemporal events. The term 'event' refers to significant changes or occurrences of particular patterns formed by consecutive attribute values [24].

We take a step further: we propose an approach called COPE (Co-Occurrence Pattern Exploration) that not only supports finding the locations and times of various events but also enables the analyst to explore what happened at other locations simultaneously with, before, or after the identified events. The ultimate goal is to support the discovery of *event co-occurrence patterns*, i.e., frequent occurrences of a specific temporal relation between events happening at two locations, such as 'same time', 'before', or 'after' (by a certain number of time steps). Please note that the term *co-occurrence* does not refer only to 'same time' cases but is used here in a more general sense.

The task of finding co-occurrence patterns is pervasive in analyzing spatial time series. Let us take economy data as an example. When the economy of a region develops rapidly, it may stimulate economic growth in other regions, but it may also have an opposite effect by attracting investments and talents from other regions. An analyst may wish to find which of the two possible cases actually happened, when, and where. This may give important clues to understanding economic problems.

Since the notion of relevant event is formed in the mind of the analyst in the process of data exploration and may change over time, flexible event definition and extraction need to be supported. Moreover, the analyst may wish to look for diverse patterns in terms of the target location where the events occurred, the kind of the accompanying events (e.g., same as or opposite to the events at the target location), and the temporal relation between the events at the target location and the accompanying events at this and other locations.

The analytical process thus consists of two stages: 1) define and identify relevant events and 2) find co-occurring events and explore the temporal distribution and the strengths of the co-occurrence relationships for different pairs of locations. Our goal is to support the whole process in such a way that the analyst has high flexibility in choosing and changing the analysis focus.

- *J. Li is with the School of Computer Software, Tianjin University, China. E-mail: jie.li@tju.edu.cn.*
- *K. Zhang is with the University of Texas, Dallas, TX. E-mail: kzhang@utdallas.edu.*
- *S. Chen is with the Fraunhofer Institute IAIS, Rhein-Sieg, North Rhine-Westphalia 53757, Germany. E-mail: siming.chen@iais.fraunhofer.de.*
- *G. Andrienko and N. Andrienko are with the Fraunhofer Institute IAIS, Rhein-Sieg, North Rhine-Westphalia 53757, Germany, and also with the City University, London EC1V 0HB, UK. E-mail: {gennady.andrienko, natalia.andrienko}@iais.fraunhofer.de.*

Manuscript received 24 Nov. 2017; revised 18 June 2018; accepted 22 June 2018. Date of publication 29 June 2018; date of current version 28 June 2019. (Coressponding author: Jie Li.)
Recommended for acceptance by H. Theisel.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2018.2851227

The proposed framework seamlessly supports the stages of event extraction and co-occurrence exploration. It makes the following contributions:

1. *A general definition of co-occurrence pattern.* We model the co-occurrence pattern as a temporal relation between two event sets and propose a method for quantitation of the pattern strength. The definition provides a foundation for the analysis.
2. *An analysis framework for detecting and exploring co-occurrence patterns in spatial time series.* The framework enables interactive event extraction and flexible co-occurrence exploration.
3. *A visual analytics tool that implements the framework.* It integrates three novel visualization components and allows the analyst to move seamlessly between the two stages in the analysis, event extraction and co-occurrence exploration.

The main *novelty* of our work is tackling an analytical problem that, to the best of our knowledge, was not yet addressed in the visualization and visual analytics research, as will be shown in the next section.

The remaining part of this paper is organized as follows. Section 2 reviews related work. Section 3 defines the problem we are striving to solve and the requirements to be fulfilled. The proposed solution is described in Section 4. A visual design of the tool that implements the framework is presented in Section 5. Section 6 describes two analytical case studies, followed by a discussion in Section 7. We conclude the paper in Section 8.

2 Related Work

We review the related work regarding the four aspects: analysis of spatial time series and events, event extraction from time series, spatiotemporal visualization, and analysis of event co-occurrences.

2.1 Analysis of Spatial Time Series and Events

Spatial time series is an important type of data in many fields [12], such as, economy [25], global trade [27], traffic [35] and emergency response [33]. Research has been focusing on situation understanding [9], forecasting [29], anomaly detection [28], and relationship identifying among different parts of data [59]. The proposed methods have acknowledged limitations [23]; thus, they lack the capability to display the results in a form that is easily perceived by humans and to involve human analysts in the analysis process. The visual analytics methods dealing with spatial time series [8], [36], [43] mostly focus on summarizing the data and providing comprehensible overviews rather than supporting more detailed analyses needed for detection of various events and exploration of relationships between them.

There exist visual analytics approaches supporting analyses of event sequences. Unger et al. [51] find categorical event sequences with high semantic and temporal similarities. Cappers and Wijk [14] detect event sequences satisfying user-defined rules. Du et al. [21] find past event sequences that are partly similar to users' event sequences and lead to users' desired goals. The problems that are tackled in all these works differ from ours. Our focus is temporal relationships between events from different sequences rather than similarity of the sequences. Monroe et al. [40] provide an aggregated display of a large number of event sequences and enable detecting occurrences of same events or subsequences at the same relative positions in multiple event sequences. Our work differs in several respects: (1) we deal with a variety of temporal relationships, not only 'same time'; (2) we focus on discovering not just instances but *patterns* of co-occurrence, when multiple such instances exist for sequence pairs; (3) as our events occur in space, we enable exploring their spatial locations and distributions.

2.2 Event Extraction from Spatial Time Series

Extracting goal-relevant events from original data is an effective way to reduce data volume and pattern variety to sharpen analytical focus [21]. Many researchers focused on detection of peaks or pits in the time series. Andrienko et al. [4] proposed a peak detection algorithm. Chen et al. [18] defined a peak model of a MOOC video clickstream and used it to find the sequences with frequent operations. Gregory and Shneiderman [24] developed a visualization system allowing the user to identify different types of shapes in time series: spike, sink, rise, drop, plateau, valley, and gap; such shapes are called 'motifs' in data mining and statistics [15], where much research on event detection has been conducted [26], [56]. A common feature of the event detection algorithms and systems is their focusing on particular shapes irrespectively of the values of the attributes, while the latter may be quite important in analyzing specific data. Our framework allows the analyst to define events of interest in a flexible way based on the absolute attribute values or the character and rate of change.

2.3 Spatiotemporal Visualization

The combination of space, time, and multiple attributes creates such information richness and heterogeneity that makes spatial time series impossible to visualize using standard charts. The possibilities for a combined visualization of the spatial and temporal aspects of data in a single display are very limited, and the existing scarce approaches are not scalable to large amounts of data [7]. A common approach is to use coordinated views [44], [32], where map displays show the spatial aspect and various kinds of time-oriented displays [1], [3] show the temporal aspect. Waldner et al. [53] used curves to link multiple aspects of objects in different projection space. Andrienko et al. [6] and Landesberger et al. [31] resorted to the spatiotemporal aggregation and clustering respectively. Tominski et al. [50] proposed a stacked visualization technique in 3D space. Elzen et al. [22] and Bach et al. [11] sequentially connected points in a projection obtained by means of dimensionality reduction to show temporal patterns of evolving multi-dimensional features. Different from the existing works, we utilize a small-multiple method, where multiple maps are arranged horizontally according to time and vertically according to intervals of attribute values. Various interactions enable exploring a large number of objects over a long time period.

2.4 Co-occurrence Analysis of Spatial Events

A lot of research related to co-occurrence analysis is done in data mining, where it is acknowledged that the

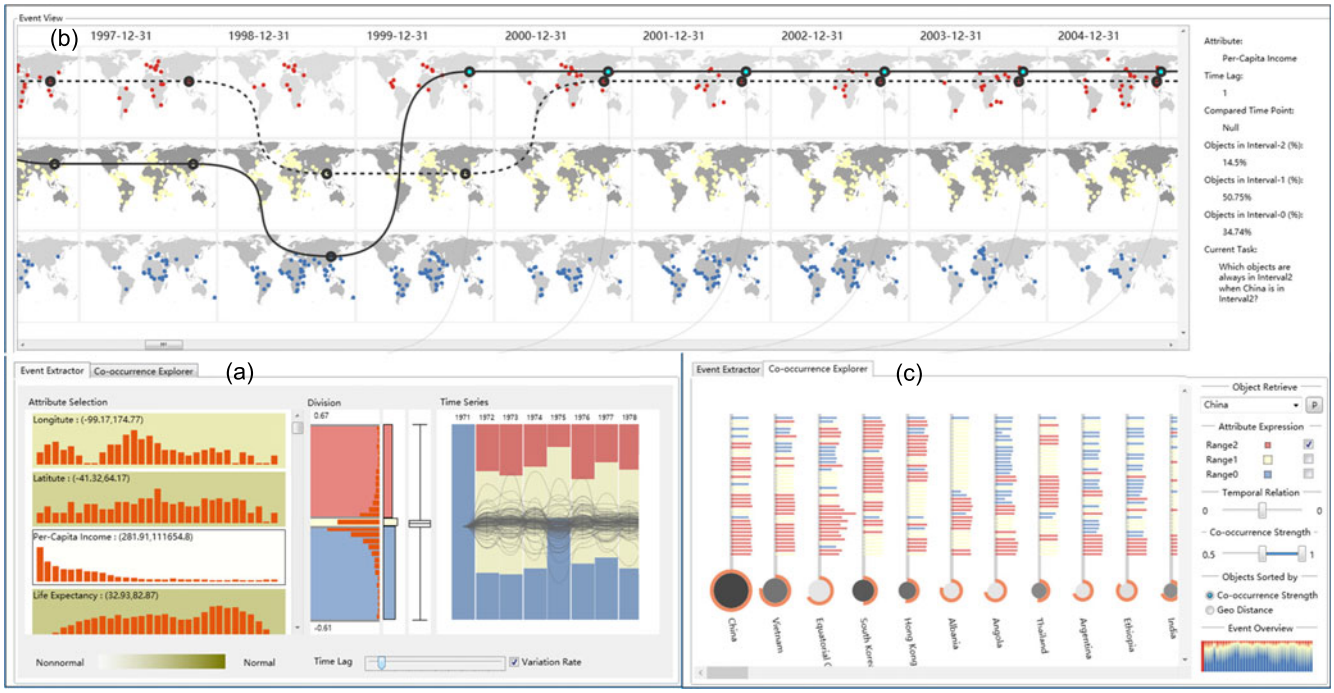

Fig. 1. COPE is a visual analytics framework consisting of three components: (a) Event Extractor, which extracts interactively defined events of interest from spatial time series, (b) Event View, which shows the distribution of the extracted events over space and time, and (c) Co-occurrence Explorer for interactive exploration of co-occurrence patterns. In this example, the user is exploring which countries had high per capita income growth rates at the same times as this happened in China.

spatiotemporal context of events increases the analysis complexity [10]. A related earlier concept is colocation [46], which means that objects are frequently located close to each other [57], [58]. Later, researchers began to use the concept of co-occurrence to consider colocation phenomenon within semantically rich spatiotemporal context. Numerous specific types of spatiotemporal co-occurrence patterns with different spatial and temporal constraints have been introduced [16], [17], [42], [60], which reflects the variety of co-occurrence patterns.

Exploring co-occurrence patterns of predefined spatial events has also received attention in the visualization community [13], [37], [55]. The task of event definition and detection was not addressed in these works. Most of them aim at finding individual co-occurrences of events rather than persistent constant co-occurrence patterns. The work by Peuquet et al. [41] focuses on patterns at a high spatial scale rather than relationships between specific locations. Wu et al. [55] analyze co-location phenomena based on human movement trajectories. Chen et al. [19] use geolocated social media data to discover links between locations in terms of people movement between them. Their goal is different from detecting relationships based on event co-occurrences, which is the focus of our work.

Our review shows that the problem of flexible interactive event extraction from spatial time series and discovery of location-specific patterns of event co-occurrence has not been addressed yet in visual analytics.

3 Problem Statement

In this section, we define the concept of co-occurrence pattern and set the requirements to guide the design of the analytical framework.

3.1 Definition of Co-Occurrence Pattern

To define the concept of co-occurrence pattern, we need to introduce a few other definitions.

Definition 1. *Let* $L = \{l_1, l_2, \ldots, l_M\}$ *be a finite set of spatial locations and* $T = (t_1, t_2, \ldots, t_N)$ *is a sequence of consecutive time steps, where 'step' may refer to a moment or interval in time. Let* A *be a set of attributes characterizing the locations at different times, i.e., for each location* l_i *there is a temporally ordered sequence* $(a_{i1}, a_{i2}, \ldots, a_{iN})$*, in which every* a_{ij} *is a combination of attribute values characterizing the location* l_i *at time step* t_j*. Each such sequence is called* **location-specific time series***, or, shortly,* **local time series***. The set of local time series for all locations in* L *is called* **spatial time series***.*

The illustration in Fig. 2 shows that spatial time series can be represented as a matrix with the two dimensions corresponding to the locations L and time steps T, the cells being filled with the attribute values or combinations. The latter case (i.e., when there are multiple attributes) can also be represented as a cube where the third dimension accommodates the different attributes and the cells contain values of singular attributes.

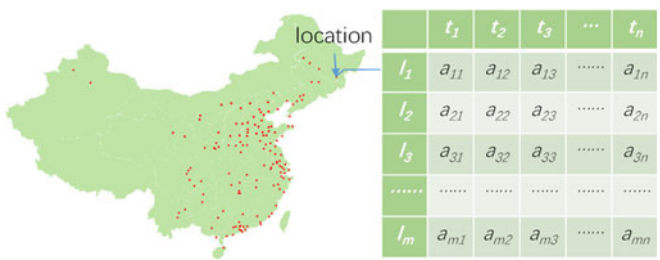

Fig. 2. Illustration of the spatial time series. Each red circle represents a city in the China air quality observation system.

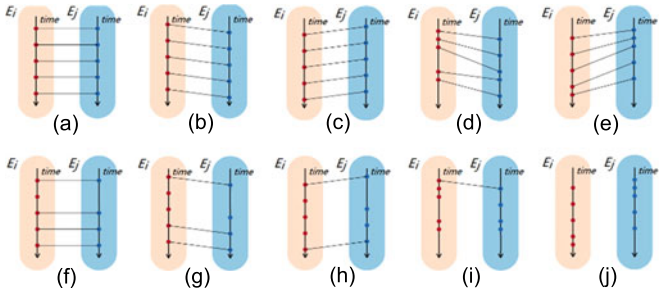

Fig. 3. Illustration of co-occurrence patterns between sets of events E_i and E_j that occurred at locations l_i and l_j. Pink and blue shapes represent two event sets, each containing five events marked with red and blue points. (a-e) Five types of temporal relations. (f-j) Co-occurrence patterns with different strengths.

The concept of the spatial event has been defined as any physical or abstract entity that only exists at some moment or a limited interval in time and has a location in space [5]. According to this definition, any element a_{ij} of a spatial time series is a spatial event (it exists at location l_i and time moment or interval t_j) and, moreover, any subsequence a_{ij}, $a_{ij+1}, \ldots, a_{ij+k}$ is also a spatial event (it exists at location l_i during time interval $[t_j, t_{j+k}]$). However, not all these events may be relevant to the goal of analysis but only events with particular properties. The properties that distinguish relevant events from the rest of the data can be, in general, specified by a predicate $P(A)$, which includes one or more conditions concerning the attribute values. A condition may refer either to the values themselves or to changes of the values compared to the previous time step or another time step in the past, or over a time interval. For example, a condition may specify that the growth rate of per capita income must be over 5 percent.

Definition 2. *A* **relevant event**, *also called* **event of interest**, *with respect to a given relevance predicate* P *is a tuple* $\langle l_i, t_j, a_{ij}, P\rangle$ *such that* $P(a_{ij}) = true$.

Hence, a relevance predicate defines a set of relevant events that can be extracted from spatial time series. We shall use the notation $E(P)$ for a set of relevant events according to a predicate P and denote a member event from this set as e^P, i.e., $e^P \in E(P)$. A relevant event that occurred at location l_i and time t_j will be denoted $e^P(l_i, t_j)$. The subset of all relevant events that occurred at location l_i will be denoted as $E_i(P)$. The whole set $E(P)$ is a union of the subsets of relevant events that occurred at all location $l_1, l_2, \ldots, l_M$: $E(P) = E_1(P) \cup E_2(P) \cup \ldots \cup E_M(P)$.

Definition 3. *Let* $E_i(P)$ *and* $E_j(Q)$ *be sets of relevant events according to predicates* P *and* Q *that occurred at locations* l_i *and* l_j, *respectively. Let* R *be a specific temporal relation, such as 'same time', 'before', or 'after' [2]. A* **co-occurrence pattern** *between locations* l_i *and* l_j *is the set consisting of all pairs of relevant events* $(e^P(l_i, t_x), e^Q(l_j, t_y))$ *such that* $e^P(l_i, t_x) \in E_i(P)$, $e^Q(l_j, t_y) \in E_j(Q)$, *and the expression* $t_x\, R\, t_y$ *is true, i.e., the specified temporal relation* R *holds between the times of the event occurrence.*

This definition allows P and Q to be the same predicate ($P = Q$); obviously, in this case, $E(P) = E(Q)$.

The upper part of Fig. 3 schematically illustrates different temporal relations that can exist between relevant events from two subsets E_i and E_j that occurred at locations l_i and l_j, respectively: at the same time (a), earlier or later by a fixed temporal interval (b-c), and earlier or later within a given maximal temporal interval (d-e).

A ***co-occurrence pattern*** between two locations l_i and l_j can be denoted as $C_{ij} = (l_i, l_j, P, Q, R)$, where P and Q are relevance predicates and R is a temporal relation.

Definition 4. *The* **cardinality** *of a co-occurrence pattern* C_{ij}, *denoted* $C_{ij}|$, *is the number of event pairs in it.*

Definition 5. *The* **strength** *of a co-occurrence pattern* C_{ij} *for* l_i *is the ratio between the cardinality of the pattern and the cardinality of the subset* E_i *of relevant events that occurred at location* l_i:

$$Si \;=\; |Cij\,|/|Ei\,|.$$

The value of s_i ranges in [0, 1], where $s_i = 1$ means that the events of E_j always co-occurred with some events of E_i, while $s_i = 0$ means that they never co-occurred. Figs. 3f, 3g, 3h, 3i, and 3j) schematically illustrates examples of co-occurrence patterns with $s_i = 0.8,\ 0.6,\ 0.4,\ 0.2,\ 0$ ($s_j = 1,\ 0.75,\ 0.5,\ 0.25,\ 0$). Note that the co-occurrence strength S_i and S_j are different when the number of events in E_j and E_j are different. The ***goal of co-occurrence pattern discovery*** is to find co-occurrence patterns whose strengths and cardinalities are sufficiently high.

3.2 A Simplified Approach to Defining Event Relevance

The conceptual model presented in the previous subsection involves a very general way of specifying which events are relevant, namely, by creating an arbitrary predicate concerning attribute values. While this allows very high flexibility in defining relevance, constructing predicates may be a difficult and time-consuming task for the analyst. After consulting with several experts from different domains and studying examples of time series analysis available in the literature, we came to a conclusion that analysts are most often interested in detecting two kinds of events: 1) occurrence of very high or very low values and 2) value change, such as increase or decrease, possibly, happening with a certain minimal rate. Regarding event co-occurrences between locations, analysts are often interested in either the same kind of events or *opposite* events, such as economic growth and decline, occurring at different locations.

Accordingly, the approach to defining relevance can be substantially simplified. Let us first consider the case of analyzing spatial time series with a single numeric attribute A. Let $[a_{min}, a_{max}]$ be the value range of A and $D(A)$ be a division of this range into three subranges $[a_{min}, a_{low}] + [a_{low}, a_{high}] +]a_{high}, a_{max}]$, where are a_{low} and a_{high} are two chosen threshold values such that $a_{min} < a_{low} \leq a_{high} < a_{max}$. This division is suitable for defining simultaneously three kinds of relevant events: occurrence of low values, high values, and medium values. The threshold values a_{low} and a_{high} can be interactively set by the user. Selection of one of the three value subranges implicitly creates a relevance predicate, which is an easy way to define relevant events.

This approach can be extended to relevant events in terms of value changes by involving appropriate transformations

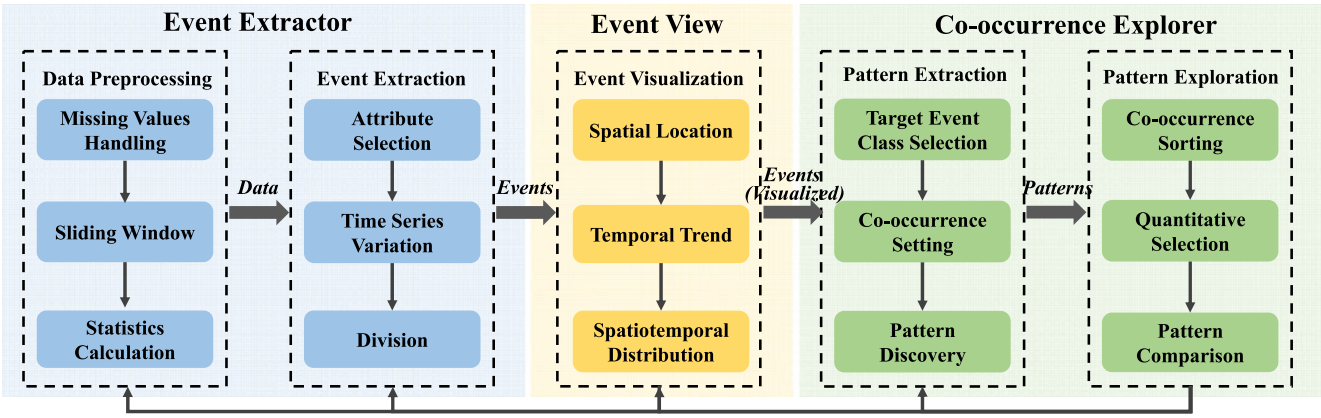

Fig. 4. Analytical Pipeline. The pipeline enables drilling down from original data to pattern discovery. It is supported by three interactive visual components marked with different background colors.

of the data. Thus, the original values can be transformed to differences or ratios with respect to the previous time steps or, more generally, to time steps $t_i - \Delta t$, where t_i is the current step and Δt is a chosen temporal lag. Another possible transformation is computing the difference or ratio to the mean or maximal value from the preceding time interval of a chosen length. Let A' be the derived attribute resulting from the transformation. By specifying a division $D(A')$ in a way similar to $D(A)$, the analyst can define relevant events of increase, decrease, and absence of significant change. Moreover, the threshold values can specify the minimal amplitudes or rates of increase and decrease that are of potential interest. To define such events, the analyst can choose a suitable transformation from a list. After the transformation is applied, the analyst interactively sets the thresholds. Again, this way of event definition is easy for the analyst.

The idea of division can also be extended to values of qualitative (categorical) attributes and to value combinations of multiple attributes; the analyst partitions a set of all values or combinations into three subsets such that two of them can be considered as opposite to each other and the third includes all the rest. This is quite easy to do for a single qualitative attribute. Dealing with multiple attributes requires more effort. The analyst may be proposed to use a parallel coordinate plot or a spider diagram the axes of which are suitably oriented and scaled (e.g., according to value quantiles), enabling convenient selection of combinations of high values of some attributes and low values of other attributes.

In our prototype implementation, we enable defining relevant events based on a single numeric attribute, which is sufficient for a proof of concept.

3.3 Analysis Tasks

Our analytical framework has been designed to support the following analysis tasks:

T1. Extract all relevant events and see the locations and times of their occurrence.

T2. For a chosen target location, see the times of event occurrence at this location and the spatial patterns of other events occurring at the same times, before, and after.

T3. For a chosen target location, find other locations where events frequently occur simultaneously or with a specific time difference to the events at the target location. Assess and compare the strengths of these co-occurrence patterns between the locations.

T4. For a chosen pair of locations, see the cases when the specified temporal relationships between events occurrences at these locations do and do not take place.

3.4 Functional Requirements

According to the specific nature of spatial time series and the analysis tasks being addressed, we have identified the following functional requirements for supporting event co-occurrence analysis:

R1: Flexible definition and extraction of relevant events. As we discussed earlier, the notion of the relevant event may not be predefined in the mind of the analyst but may be formed after initial data exploration and evolve in the course of further analysis. Hence, the analyst should be able to flexibly define and re-define the relevant events at any time during the analysis and immediately see the resulting set of extracted events.

R2: Allowing for diverse co-occurrence relationships. The analyst should be able to find co-occurrence patterns involving the same or opposite kinds of events for different target locations and temporal relations.

R3: Interactive spatiotemporal visualization capable to accommodate long time series. The analyst must be able to see the overall distribution of the extracted relevant events over space and time and, when desired, focus on neighborhoods of interesting locations and/or on particular time intervals.

4 Analytical Framework

The general analytical pipeline for event co-occurrence analysis is shown in Fig. 4. It works in an iterative process, in which the analysts can return to any of the previous steps, e.g., to modify the relevance predicates, choose another attribute, or transform the data.

Data Preprocessing. The pre-processing (Fig. 5a) may include handling of missing values, data aggregation by suitable time steps (e.g., hourly measurements into daily summaries) and/or areas, data smoothing for reducing the noise, and calculation of overall statistical criteria, such as normality, for each attribute. The criteria show the general features of the attributes, thus helping the analyst to

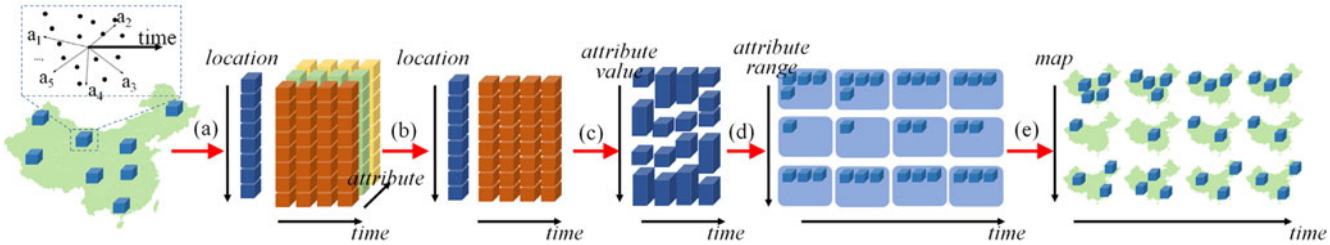

Fig. 5. Illustration of event extraction and visualization. (a) Data preprocessing. (b) Attribute selection. (c) Time series variation. (d) Division. (e) Event Visualization. Blue cubes represent objects having fixed spatial coordinates, and colorful cubes represent attributes.

determine which attributes may have distinctive co-occurrence patterns.

Event Extraction. This stage includes choosing an attribute to be used for relevant event definition (Fig. 5b), adjusting time series of the selected attribute (Fig. 5c), and division of its value range (Fig. 5d), as explained in Section 3.2. In response, the corresponding set of relevant events is extracted from the data.

Event Visualization. The extracted relevant events are visualized so that their spatial locations and temporal references could be easily identified (Fig. 5e). The analyst is able to focus on all events that occurred at a chosen target location during the whole time period.

Pattern Extraction. The analyst selects the attribute value intervals specifying the relevance predicates P and Q (in particular, they may be the same); see the definitions in Section 3.1. The analyst also specifies the temporal relation that must hold between the times of co-occurring events and chooses a target location. The co-occurrence patterns are extracted based on Definition 3, and their strengths are determined (Definition 5).

Pattern Exploration. Analysts interactively explore the extracted co-occurrence patterns from different perspectives. For example, they can sort the *patterns* associated with different locations to find which locations have stronger co-occurrence patterns with the target location and how far they are located in space with respect to the target location.

The pipeline is supported by a component architecture shown in Fig. 4. It consists of three tightly interrelated components: (1) *Event Extractor* that implements the first two steps of the pipeline, enabling interactive extraction of relevant events from original spatial time series. (2) *Event View* visualizes the extracted event within the spatiotemporal context. (3) *Co-occurrence Explorer* in which users can explore and understand co-occurrence patterns from different perspectives through multiple interactive operations. Of the three components, the event view provides the primary spatiotemporal context, which is also important for the other two components.

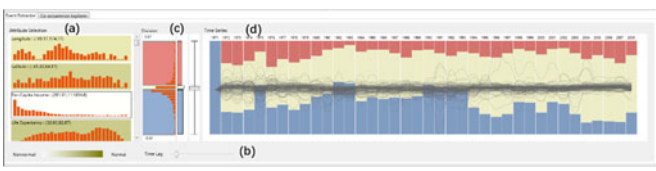

Fig. 6. Event extractor. (a) Attribute histograms for selecting an attribute based on the value distribution properties. (b) A slider and a checkbox used to transform the original values into differences. (c) A value range division control consisting of two vertical histograms and a boxplot. (d) A time series view showing the division results.

5 Visual Design

The main interface of COPE is shown in Fig. 8. The event view, which is the core visual component, is placed on the top left of the interface. Since there is no need to use the event extractor and the co-occurrence explorer simultaneously, the two components are put into a tab control at the bottom of the interface. The user can switch between the two components by clicking the corresponding tabs. This section describes the visual designs of the three visual components in detail.

5.1 Event Extractor

According to the requirement R1, the event extractor enables flexible definition and extraction of relevant events. It implements the simplified approach to creating relevance predicates as explained in Section 3.2. The current prototype implementation supports creating predicates based on a single numeric attribute. The component combines multiple visualization techniques shown in Fig. 6.

On the left (Fig. 6a), there is a scrollable list containing frequency histograms showing the value distributions of all attributes. The display reflects results of data preprocessing. The background of each histogram is colored according to the normality of the attribute. The analyst can select an interesting attribute according to the data distribution (the histogram of the selected attribute is enclosed in a rectangular frame) and use a slider (Fig. 6b) to transform the time series. If the slider value is N, the time series is transformed to represent the changes relative to the Nth time step before each step. In this case, the attribute values for all locations at the first N time steps are set to zero, because they do not have previous records for comparison. We also put a checkbox at the right side of the slider to support two transformation options, 1) the absolute difference to the previous value (the checkbox is not checked); 2) the difference in percentage to the previous values (the checkbox is checked).

After selecting an attribute, its value distribution is shown by two frequency histograms and a box plot in the division control (Fig. 6c). The two histograms are oriented vertically and have a common axis corresponding to the value range of the attribute. In the histogram positioned on the left of the axis, the bars correspond to equal value intervals of the attribute, analogously to the histograms in Fig. 6a. The histogram positioned on the right of the axis has three bars corresponding to the current division of the value range by two thresholds a_{low} and a_{high}, as explained in Section 3.2. The bars of the right histogram are painted in blue, yellow, and red colors corresponding to the low, medium, and high attribute values, respectively. The same colors are used to paint the background of the left

histogram; hence, the background represents the current division. The threshold values can be changed by dragging the lower edge of the red rectangle and the upper edge of the blue rectangle. The two histograms will also change when applying a transformation to the time series.

The result of the division is also shown in the time series view (Fig. 6d) with the time mapped onto the X-axis and the original or transformed attribute value range onto the Y-axis. The gray curves represent the time series of all locations. In the background, there is a sequence of segmented vertical bars each corresponding to one time step. The bars are divided into red, yellow, and blue segments showing the proportions of attribute values attained at each time step that fit into the low, medium, and high value subranges. By observing the sizes of the segments, the analyst can adjust the thresholds depending on whether he/she is interested in examining fewer events with more extreme attribute values or a larger number of events with values that were attained more frequently.

5.2 Event View

The event view, as in Fig. 1b, is the core visual component of our approach providing a spatiotemporal context to accommodate all the extracted events for further co-occurrence exploration. The event view has a tabular structure, which is consistent with the structure of the time series view (Fig. 6d). The columns and rows of the table correspond to the bars and the bar segments of the time series view. It means that each column corresponds to a time step and each row to one of the value subranges of the selected attribute resulting from the division. The top row corresponds to the high values, the bottom row to the low values, and the middle row to the medium values. Each cell of the table contains a map on which the locations whose values lie within the corresponding subrange are represented by dots. This enables the analyst to observe the geographical positions of the events defined by the current attribute range division. The dots are painted in red (top row), yellow (middle row), and blue (bottom row), consistently with the color encoding in the event extractor (Fig. 6). The gray shading of the background maps encode the numbers of objects shown on the maps. The darker a map is, the more objects are on it. We use a grayscale to reduce the visual clutter.

When the time series is long (i.e., has many time steps), some of the columns can be folded by clicking on the corresponding time labels; as a result, the remaining columns can be better seen. The gray shading of the background maps is not used within the folded columns. For example, in Fig. 8, the years between 1986 and 2004 are folded. The map size can also be adjusted through mouse wheel operation.

An information panel on the right of the map view shows the currently selected attribute, the time lag for computing the value change, the proportion of the events in each row, and the current analysis task performed with the co-occurrence explorer, when the latter is active.

By clicking on a dot representing an event in a map, the user selects one of the locations to be the target location. Let $l*$ be the target location and $E*$ be all events that occurred at this location, including the events with the high, medium, and low attribute values. After the selection of $l*$, a curve is drawn to connect all events from $E*$ in the chronological order across the event view. This curve is further referred to as *trend line*. COPE highlights the locations of the events on different maps along the trend line (see the big black circles on the trend line in Fig. 8).

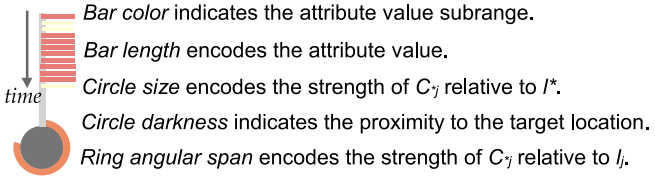

Fig. 7. The glyph design for representing a co-occurrence pattern.

Depending on whether the selection of $l*$ was done in the top, middle, or bottom row of the event view, a relevance predicate P (Definition 2) is created that selects the high, medium, or low values of the attribute, respectively. All events shown in the selected row make the set $E(P)$ of relevant events according to P. Those events from $E(P)$ that occurred at the selected target location $l*$ will be called the *target events*. The set of target events may be denoted as $E*(P)$: $E*(P) = E* \cap E(P)$.

5.3 Co-Occurrence Explorer

The co-occurrence explorer (Fig. 1c) contains a co-occurrence list representing visually the characteristics of the extracted co-occurrence patterns C_{*j} between the target location $l*$ and each location $l_j \in L$, including also the target location itself. The co-occurrence explorer also includes a set of subcomponents for controlling pattern extraction and exploration process (Figs. 8d, 8e, 8f, 8g, 8h, and 8i). Particularly, one of the controls (Fig. 8e) allows the analyst to specify the second relevance predicate Q (Definition 3). For this purpose, the analyst just needs to select the high, medium, or low subrange of the attribute values. Accordingly, the set of relevant events $E(Q)$ is selected; $E(Q) = E_1(Q) \cup E_2(Q), \ldots, E_m(Q)$, where $E_j(Q)$ is the subset of relevant events that occurred at the location l_j. Another control allows choosing the temporal relation R, as in Fig. 8f. For each location l_j, the co-occurrence pattern C_{*j} is generated according to Definition 3. That is, the pattern includes every event $e \in E_j(Q)$ for which $\exists e* \in E*(P)$ such that $e\ R\ e* = true$, i.e., there exists some event $e*$ in the target event set $E*(P)$ that e has the temporal relation R to the event $e*$.

5.3.1 Co-Occurrence List

The co-occurrence list contains a horizontally arranged set of glyphs each representing the co-occurrence pattern C_{*j} for one of the locations l_j, $1 \leq j \leq m$. A glyph consists of simple components involving commonly used visualization techniques: bar chart, proportional symbol size, and color coding (Fig. 7).

The size of the circle in a glyph represents s_*, i.e., the strength of the pattern C_{*j} relative to $l*$. The angular size of the ring segment attached to the circle shows $s_j = |C_{*j}|/|E*(P)|$, i.e., the strength of the pattern C_{*j} relative to l_j (Definition 5). Simultaneously showing the s_* and s_j can better represent the co-occurrence patterns. Thus, a bigger s_* and a smaller s_j means $|E_j(Q)| >> |E*(P)|$, i.e., the higher s_* may be occasionally caused by the frequent occurrence of $E_j(Q)$, while bigger s_* and s_j reflect stronger co-

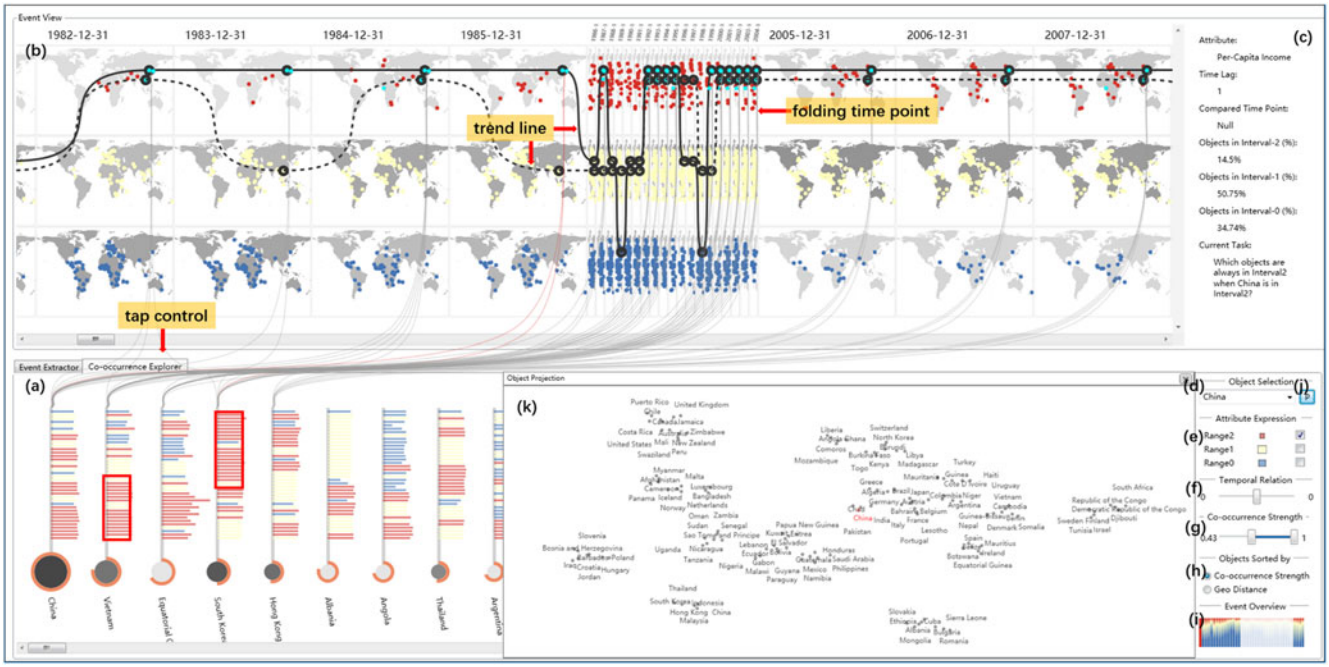

Fig. 8. Exploring which countries had a high per capita income growth rate when China's per capita income increased quickly. The countries are sorted according to the distances between them and China. (a-k) are the subcomponents of the three visual components.

occurrence relationships between l_j and $l*$. The grayscale shading of a circle encodes the geographic distance between the location l_j and the target location $l*$; darker shades mean smaller distances.

In the bar chart, the vertical axis represents time, and each bar corresponds to a time step. The glyph in Fig. 7 shows 12 time steps. The length of a bar represents the attribute value attained at the location l_j at the corresponding time step, while the color indicates its value subrange. This design provides more detailed context information concerning each location l_j.

The co-occurrence list contains a glyph for the target location $l*$. When the chosen temporal relation R is 'same time' and $Q = P$, the circle size in the glyph will be the biggest among all. The circle is always the darkest since in this case the distance to the target location is zero (see the glyph of China in Fig. 8).

5.3.2 Interactive Operations

The interaction controls support the retrieval and exploration of event co-occurrence patterns.

Selecting a target location. COPE supports three ways for selecting an initial target location. First, the analyst can directly select a spatial object in the event view. Second, the analyst can enter an object's name in a search box (Fig. 8d). Third, the analyst can select locations from a projection view (Fig. 8k), which is triggered by clicking a button next to the search box (Fig. 8j). The locations are projected according to the similarity of their event time series by means of dimensionality reduction. We use the T-SNE method, but other techniques can also be used. Collision detection is utilized to avoid overlaps in the projection. The projection provides an overview of all locations regarding the extracted events. The analyst can apply different strategies for selecting the target location, e.g., take one from a large or small cluster, or take a location that is most dissimilar to the others. The currently selected location is highlighted in the projection, its name is shown in the search box, and COPE generates its trend line to connect the positions of the location along the time.

Selecting the second relevance predicate Q. The second predicate may be the same as or different from P, which is the predicate selecting the target event set. According to the approach explained in Section 3.2, the analyst selects either the same value subrange as for P (i.e., $Q = P$) or one of the two other subranges. To enable this, COPE provides three checkboxes, each corresponding to one subrange. For example, in Fig. 8e, the first checkbox is selected. The corresponding predicate selects the high values of the attribute.

Setting the temporal relation. The temporal relation R can be set using a range slider (Fig. 1c). Each tick of the slider corresponds to a time step. The range slider has two bars, which have three usages that implement all the five types of temporal relations (Fig. 3): 1) two bars are set to zero (default setting) to select the relation 'same time', 2) two bars are set at the same tick to specify relationships 'before by Δt steps' or 'after by Δt steps', and 3) two bars are set to different values to generate a temporal interval for selecting the relations 'before ' or 'after', as in Fig. 1.

Setting a threshold for the co-occurrence strength. COPE provides a range slider for quantitatively selecting co-occurrence patterns with specific strengths, as in Fig. 8g. The two bars of the range slider represents the minimum and maximum strength thresholds. For each location having the strength of the co-occurrence pattern C_{*j} within $[minimum,\ maximum]$, the member events of the pattern C_{*j} are connected to the glyph of l_j by curved lines. This allows the user to explore the details of strong co-occurrence patterns. The lines are bundled to reduce the clutter in the event view.

Glyph sorting. COPE provides two choice buttons to select the way of sorting the glyphs, as in Fig. 8h. By default, the

glyphs are sorted according to the strengths of the co-occurrence patterns. They can also be sorted according to the distances of the respective locations to the target location $l*$. This arrangement can reveal anomalies when distant objects have strong co-occurrence patterns, or nearby objects have very weak co-occurrences.

Event overview. This subcomponent (Fig. 8i) is a miniature of the time series view in Fig. 6d; the original view is hidden in the pattern exploration mode. Besides the division into high, medium, and low attribute values, this subcomponent also shows for each time step whether it is folded in the event view. The bars for the folded steps have unsaturated coloring. The component can also be used for folding and unfolding time steps by mouse clicking or dragging.

6 Evaluation

This section reports our evaluation of the effectiveness of COPE through two case studies on different spatiotemporal scales and an expert review. As we cannot describe the case studies in full detail, we just give a few examples of questions that can be answered using COPE and show how the answers can be obtained.

6.1 Economic Statistical Data

This dataset contains long-term statistic records (1800-2008) of three economic criteria, namely, per capita income, life expectancy, and population, of 199 countries (http://bost.ocks.org/mike/nations/). We only select the records after 1970, in which most countries have relatively complete records.

Q1: Which countries often had high per capita income growth rates simultaneously with China?

This question exemplifies the most basic exploration task, in which the same attribute subrange is used for the target events and the co-occurring events (i.e., the relevance predicate Q is the same as P), and the temporal relation is 'same time'. We select the attribute 'per-capita income' and transform the original values in the time series into the differences with respect to the previous time steps by setting the time lag to 1 using the slider shown in Fig. 6b. We set the upper threshold a_{high} to 5 percent to select the events of the growth rate higher than 5 percent, which puts 14.5 percent records in the top row of maps in the event view. The lower threshold a_{low} is irrelevant to this task; it can take an arbitrary value. The information concerning the task and the corresponding event selection is shown in Fig. 8c. According to the analysis goal, the upper checkbox is checked in the co-occurrence explorer, as in Fig. 8e, which selects the high-value subrange for the co-occurring events, i.e., the set $E(Q)$. The time slider is set to 0 (Fig. 8f), which selects the temporal relation 'same time', and the glyphs in the co-occurrence explorer are chosen to be sorted according to the strengths of the co-occurrence patterns (Fig. 8h). We click on the location of China in the top map corresponding to the year 1984, when China had the growth rate above 5 percent. As a result, China becomes the target location, and the target event set is extracted, i.e., the events of China having the growth rate above 5 percent. All events corresponding to China become connected by a trend line in the event view.

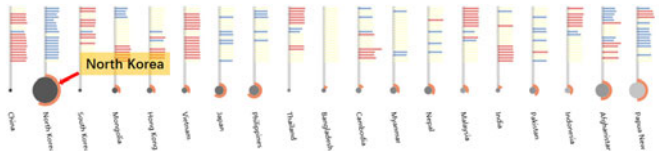

Fig. 9. Co-occurrence pattern between high growth rate in China and decline in North Korea.

The result of these selections (Fig. 8a) shows that Vietnam, Equatorial Guinea, South Korea, and Hong Kong are the top 4 countries/areas with the strongest co-occurrence patterns (more than 38 percent as in Fig. 7g). From these, Vietnam and South Korea are neighbors and major trading partners of China. By exploring the details of the co-occurrence patterns we find that the co-occurrences of South Korea mainly took place at the early stage of the period. This is opposite to Vietnam, which is more similar to China (see two red rectangles in Fig. 8a). Furthermore, we see that Vietnam has a larger angular ring span than other three countries, meaning that its rapid growths always co-occurred with those of China, and seldom occurred independently. According to the related literature [49], we know that Vietnam and China have strong economic links and mutual investments. An economic expert (Section 8) also told us that there have been many Chinese scholars studying the economic interdependence between Vietnam and China. Our findings can be used to compare Vietnam to other countries that have different co-occurrence patterns, thus finding distinguishing characteristics of the China-Vietnam relationships.

Equatorial Guinea is a special case, which is far from China but has a high co-occurrence strength. According to the literature [32], we know that the discoveries of large oil fields at the beginning of the 90's made the economic development of Equatorial Guinea rapid, which proves our finding. The expert thinks these findings are potentially very interesting, possibly leading to the discovery of implicit relationship among countries, such as teleconnection. However, the strong co-occurrence pattern may also be occasional.

Q2: Which countries often had a negative growth rate when China's per capita income increased quickly?

We set the lower threshold a_{low} to 0; thus, the low-value subrange corresponds to negative growth rates. We select this subrange using the bottom checkbox in the co-occurrence explorer. We see that the circle sizes of the most neighboring countries of China are relatively small, which shows that their per capita income was mostly growing simultaneously with China. North Korea is an exception; its circle is big, indicating a different situation in terms of the variation of the per capita income, as shown in Fig. 9. It is well known that North Korea' economy has been declining during that period, which is consistent with our finding. However, the decline is not related to the growth in China. This example demonstrates that co-occurrence patterns should be interpreted cautiously since they do not necessarily mean that the events are somehow related but only show that the events frequently co-occur.

6.2 Air Quality Observation Data of China

This dataset contains daily values of 7 pollutant attributes and 15 meteorological attributes of 196 cities during the

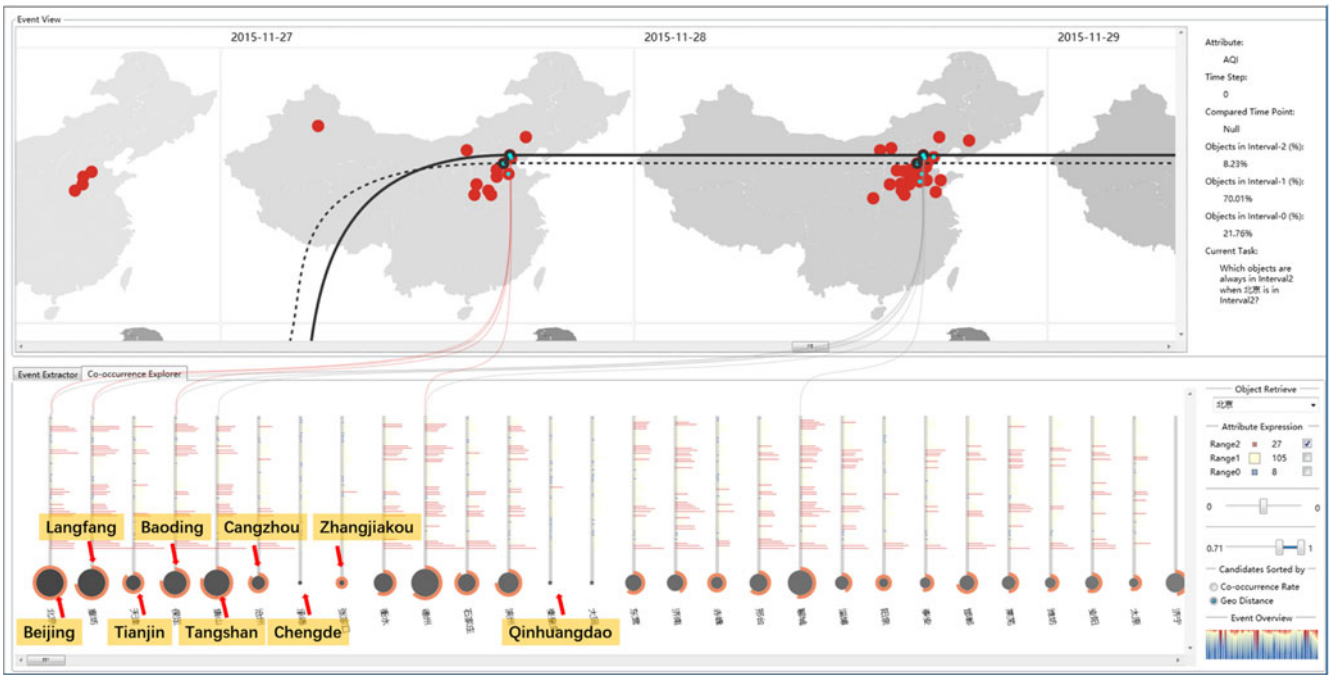

Fig. 10. Exploring which cities in China also had bad air quality when Beijing had a smog day during the period of 2015/10/1-2015/11/30, in which smog phenomena frequently occurred in China. Event view is zoomed in to clearly show the locations of Beijing's neighboring cities.

period 2015/10/01-2015/11/31. During this period, smog events frequently happened in different parts of China. Because the air quality of a city can be affected by the pollutants transmitted from nearby cities, co-occurrence pattern analysis can be helpful for finding and understanding these effects.

Q3: Which cities often have serious air pollution when Beijing has a smog day?

We choose the attribute *AQI* (Air Quality index, the higher the value is, the worse the air quality the city has). The attribute value range is divided into high, medium, and low subranges according to the China Air Quality Standard [39]. The cities shown in the top row of maps in the event view have the highest AQIs, i.e., the worst air quality. The high-value subrange is also chosen for the co-occurring events. The target location is Beijing, and the temporal relation is 'same time'. The glyphs in co-occurrence explorer are sorted according to the distances between the cities and Beijing. The map view is zoomed in to clearly observe the locations of the cities in North China, as shown in Fig. 10.

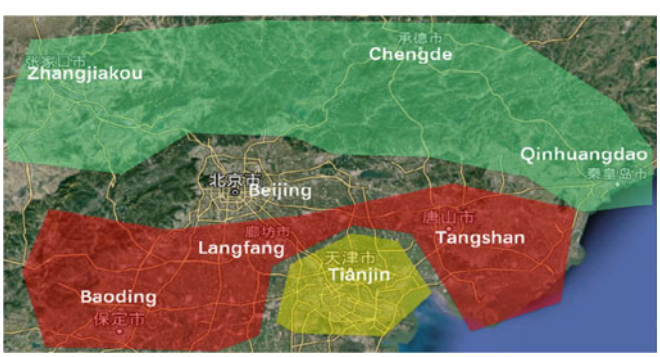

Fig. 11. Groups of the neighboring cities of Beijing. The colors of the polygons correspond to different co-occurrence strengths of the smog events in the cities to the smog events of Beijing.

The co-occurrence patterns of different cities can be compared by observing the colors and sizes of their glyphs, as in Fig 10. Three neighboring cities, Langfang, Baoding, and Tangshan, have relatively big circles, indicating that smog often occurs also in the three cities when Beijing has a smog day. The finding is consistent with the previous studies [38] which reported a strong correlation between these cities and Beijing regarding the air quality.

Tianjin and Beijing are geographically close and have similar economic developments. The size of the glyph of Tianjin, however, is not as big as the above three cities, which is consistent with our previous findings that the smog in big cities may not be as serious as in the surrounding cities [34].

We also find several cities which are close to Beijing but tend to have much better air quality than Beijing, such as Zhangjiakou, Chengde, and Qinhuangdao. According to the geographical literature [48], Zhangjiakou and Chengde have different terrains (e.g., higher altitude) and meteorological conditions (longer winter) from Beijing, resisting smog formation. Qinhuangdao is located at the seaside, which makes pollutants dissipate quickly.

We group the cities according to the strengths of their co-occurrence patterns and draw the division result on a satellite map. In Fig. 11, the red, yellow and green polygons respectively cover the cities that have high, medium, and low co-occurrence strengths. We find that, in general, smog events of Beijing always have strong co-occurrence relationships with those of the cities located in the south Beijing (see the red polygon in Fig. 11), while smog events of the cities located in the north of Beijing occur more independently (see the green polygon in Fig. 11).

Q4: In which cities does the air quality often greatly deteriorate one day before this happens in Beijing?

To answer this question, we use the time slider to choose the temporal relation 'one day before'. We also choose

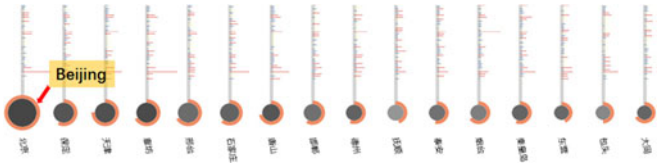

Fig. 12. Query result for Q4. The size differences among the circles are not obvious, which indicates similarly low co-occurrence strengths.

the glyphs to be sorted according to the co-occurrence strengths.

The query result is shown in Fig. 12. We find that most neighboring cities of Beijing are placed at the beginning of the list, i.e., they have the strongest co-occurrence patterns, but the size differences of the circles are not as obvious as in the previous examples. This contradicts our expectation to find the pollution sources of Beijing. We showed the result to an environmental expert. He told us this was mainly because the temporal resolution of the data is low (daily) whereas pollutants may spread to neighboring cities within a couple of hours. Hence, the available data do not allow uncovering such effects. However, he thought that our approach and techniques are interesting and valuable. At present, environment experts mainly depend on pollutant composition analysis and numeric models to analyze the smog transmission mechanism and determine the external pollutant sources of cities. Our approach being applied to high-resolution data can help them estimate a refined spatial scope by excluding the data of the cities having distinctively weaker co-occurrences, thus greatly reducing the data loads for further model computation. Furthermore, to increase the model accuracy, they can set different weights to different cities according to their co-occurrence strengths.

6.3 Expert Feedback

We conducted a small laboratory study to evaluate how easily and effectively people can use COPE to explore co-occurrence patterns, as well as to find design issues for improving the visualization and interaction design.

We recruited 12 participants, of which three were female. The ages were between 29 and 46, with the average age 36.67. To better evaluate the learnability of COPE, the invited participants were experts from several domains, including 3 HCI professors, 1 graphics associate professor, 2 economic experts, 2 meteorology researchers, 1 traffic expert, 1 project leader of multiple big data projects, 1 interface designer, and 1 medical professor. These experts did not participate in the design of COPE and thus could be seen as average users with sufficient education for understanding the nature of the data and the analysis tasks COPE had been designed to support.

We used a laptop as the experiment environment, which was sufficient for smooth visualization and interaction due to a low computation complexity of COPE.

Each session of the study started with an explanation of the usage of COPE. Then the participants were asked to select one dataset (Sections 6.1 and 6.2) and explore freely the potential co-occurrence patterns during 10-20 minutes using the experimenter's assistance when needed. The experimenter encouraged the participants to "think-aloud" and report anything they found. The participants were explicitly asked to tell any negative comments they might have. At the end of each session, the experts were asked which functions were the most useful for them.

In general, the feedback of the participants was positive. Most experts agreed that the view operations and query controls are easy to learn. They could flexibly set attribute thresholds and directly select the relevant events with default settings. In fact, COPE supports simple types of interactive operations, e.g., the user clicks on a location and views the sizes and colors of the glyphs in co-occurrence list. This was encouraging for novice users, who thus could conduct sophisticated exploration tasks with a very brief introduction and minimal assistance.

Most experts believed that COPE can accommodate data with quite a large number of spatial locations and a long period of time due to the functions of map zooming and time folding. They also affirmed that COPE was suitable for analyzing data of different domains. Five experts told that they would like to apply COPE to their data as soon as possible. For example, the traffic expert would like to analyze interrelations between roads: when the vehicle flow on a road increases or decreases, which other roads have similar or opposite variations? This may reveal issues in the road network planning. The economic expert suggested us to apply COPE to datasets with additional economic attributes. The medical professor believes that the approach of COPE may be used for analyzing epidemic spreads and, possibly, disease causes by exploring which symptoms frequently co-occur with a disease.

Three functions, i.e., trend line, location sorting, and location search were mostly appreciated by the experts and frequently used during the experiment. One HCI professor pointed out that the trend line not only clearly showed the spatiotemporal features of the events at the target location but also allowed convenient selection of the target event set. He also told that the object sorting function was useful for understanding the effects of the geographic distance on the co-occurrence patterns. The interface designer told us that the location search function was useful for her because she was not familiar with the locations of countries. She thinks that this function may also be helpful when there are too many different locations on a map making map-based selection difficult.

The experts also gave us many valuable suggestions. The meteorology researchers told us that pollutants transmission is always affected by the meteorological parameters, such as wind and air temperature. They suggested us to add a data filtering function to support exploring the co-occurrence patterns under different meteorological conditions. The project leader said that COPE may have a good commercial value. He suggested us to add several views and functions, such as data table, to form a commercial tool specialized in analyzing spatial time series. He also suggested us to develop a web version of COPE to increase its visibility and to collect the users' operation logs and suggestions to further improve its usability. An economic expert questioned our method of computing the distance between countries. COPE computes the Euclidean distance between the capitals of two countries for simplicity, whereas other distance criteria and computing methods could be more meaningful. Thus, the geometric centers could be taken instead of the capitals, and the Euclidean distance could be

replaced by the Great Circle distance on the Earth surface (which, probably, would not affect much the sorting of the glyphs in the co-occurrence explorer). According to his suggestion, we will provide more distance options for interactive selection in the future.

7 Discussion

The problem of analyzing co-occurrences between various events that may be identified in spatial time series is challenging due to countless possibilities for defining relevant events and co-occurrence relationships of interest. We have proposed an analytical pipeline to address this problem, which is supported by a visual analytics system. The case studies and expert review demonstrate the effectiveness of our approach.

Notably, our framework enables very high flexibility in setting the analysis focus and goals. It supports the discovery and exploration of patterns of *co-*, *pre-*, and *post-*occurrence of *same or opposite* events for location pairs. However, we also acknowledge some potentially problematic aspects. Below we discuss these aspects and possible ways to deal with them.

Relevance definition. The current prototype allows setting relevance predicates based on values of a single numeric attribute. However, as discussed in Section 3.2, this is not a principal limitation of our approach, since the idea of value domain division can be extended to qualitative attributes and to combinations of multiple attributes. This requires including additional interactive controls in the user interface.

Sensitivity to parameters. The relevance predicates in COPE are specified by setting exact threshold values a_{low} and a_{high}. Even small changes of these values can affect the extracted sets of relevant events $E(P)$ and $E(Q)$ and, consequently, the co-occurrence patterns that will be retrieved. COPE provides visual tools for exploring the sensitivity to the threshold settings and choosing suitable values. Specifically, the histogram and time series displays in the Event Extractor provide immediate visual feedback showing the effect of current settings on the number and temporal distribution of the extracted events.

Confirmation bias. The visualization and visual analytics communities are highly concerned with the problem of possible impacts of human cognitive biases on the course and results of the exploration and analysis [20], [54]. From the various kinds of cognitive biases [52], users of COPE may be especially prone to the confirmation bias, which means looking for patterns that confirm their pre-existing hypotheses rather than trying to discover something unexpected. This may happen, in particular, in selecting target locations. The projection display (Fig. 8k) is meant to encourage the analyst to consider the entire set of locations before selecting. The analyst can explore similarities and differences among the locations and, hopefully, find potentially interesting targets that were not previously thought of. However, this does not fully preclude the possibilities for cognitive biases, which is hardly possible at all.

Data noise. In time series analysis, it is acknowledged that data may have irregular fluctuations, that is, noise is treated as an indispensable component of time series [47]. Noise and missing values can be handled in our system using techniques for data smoothing and interpolation over a sliding temporal window. Of course, if the noise results from frequently reoccurring measurement errors, the revealed patterns can hardly be trusted. Therefore, the significant errors in data need to be detected and corrected before the analysis procedure.

Performance. In general, COPE can accommodate hundreds of objects over a long time points. For good performance, we establish various data references using hash structures. The initialization of the visual displays involves time consuming visual renderings. This process takes several seconds and may become longer as the numbers of locations and time points increase. The initialization step, however, improves the scalability of the system and enables real time responses of interactive operations in the further use of the system. Separating the computationally intensive algorithms from the business logic also minimizes the CPU and memory usage, allowing COPE to smoothly run on a laptop.

Scalability of visualization. Scalability issues are partly mitigated by the time folding, map zooming, and scroll bars in the event view and co-occurrence view. COPE, however, will be challenged when the data grows. For example, the event view can only simultaneously show a limited number of locations and time steps. To show more time steps, we have to reduce the width of the maps. However, with a large number of time steps, maps are too narrow to clearly display the spatiotemporal distribution of the events. To cope with this problem, time filtering and aggregation can be used to filter out or merge time steps without interested events or having similar event distributions.

Visual overlapping. During the analysis process, the links between the events in the event view and the glyphs in the co-occurrence explorer may cover other visual elements, and it is difficult to trace a link when too many links are drawn. In principle, the links are only used to identify locations with specified co-occurrence strengths, which are not shown by default. It is uncommon and unnecessary to simultaneously highlight too many locations. We have considered to use other visual clues instead of the link lines, such as adding a border or changing the background color of the glyphs. Link lines, however, make the co-occurrence patterns traceable in the event view, which is not supported by other visual clues. Therefore, we finally chose to use the links.

Glyph design. Although none of the experts involved in the evaluation had problems with interpreting the glyphs in the co-occurrence list, we acknowledge that the glyphs may appear complicated to average users and see a need in a further study of the glyph interpretability.

Similarity-based projection. The projection (Fig. 8k) creates a new, artificial space, in which locations are arranged based on similarities of their time series, i.e., differently from their usual arrangement on a geographic map. It may be counter-intuitive to users and thus cause interpretation difficulties. This aspect also requires investigation in further user studies.

8 Conclusion

This paper has presented a visualization approach to the exploration of co-occurrence patterns between events in spatial time series. We proposed an analytical framework based on three novel visual components for defining relevant events and extracting them from spatial time series,

visualizing the spatiotemporal distribution of the events, and exploring event co-occurrence patterns between locations. The three components are seamlessly integrated within the system COPE, whose effectiveness was tested using real-world environmental and economic data. It can be applied to spatial time series of different domains, such as the ocean, census, traffic, etc.

In the future, we plan to improve COPE in two aspects. First, we plan to support multi-attribute events, which are more common in environmental sciences. Second, we plan to address the problem of sensitivity to thresholds. We also plan to conduct a thorough empirical user study to evaluate COPE in both laboratory and practical settings.

Acknowledgments

The work is partially supported by National NSFC project (Grant number 61602340), National NSFC project (Grant number 61572348), National High-tech R&D Program (863 Grant number 215AA020506), German Priority Research Program SPP 1894 on Volunteered Geographic Information, EU project Track&Know (Grant agreement 780754), and Fraunhofer Cluster of Excellence "Cognitive Internet Technologies".

References

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*, Berlin, Germany: Springer, 2011
[2] J. F. Allen, "Maintaining knowledge about temporal intervals," *Commun. ACM*, vol. 26, no. 11, pp. 832—-843, Nov. 1983.
[3] G. Andrienko, N. Andrienko, U. Demsar, D. Dransch, and J. Dykes, "Space, time and visual analytics," *Int. J. Geographical Inf. Sci.*, vol. 24, no. 10, pp. 1577–1600, 2010.
[4] G. Andrienko, N. Andrienko, M. Mladenov, M. Mock, and C. Poelitz, "Extracting events from spatial time series," in *Proc. 14th Int. Conf. Inf. Vis.*, 2010, pp. 48–53.
[5] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel, *Visual Analytics of Movement*. Berlin, Germany: Springer, 2013.
[6] G. Andrienko, and N. Andrienko, "A general framework for using aggregation in visual exploration of movement data," *Cartographic J.*, vol. 47, no. 1, pp. 22–40, 2010.
[7] N. Andrienko, G. Andrienko, and P. Gatalsky, "Exploratory spatio-temporal visualization: an analytical review," *J. Visual Languages Comput.*, vol. 14, no. 6, pp. 503–541, 2003.
[8] P. Accorsi, N. Lalande, M. Fabrègue, A. Braud, P. Poncelet, A. Sallaberry, and F. Ber, "HydroQual: Visual analysis of river water quality," In *Proceedings of IEEE VAST'14*, pp. 123–132, 2014.
[9] E. Austin, B. A. Coull, A. Zanobetti, and P. Koutrakis, "A framework to spatially cluster air pollution monitoring sites in US based on the PM 2.5 composition," *Environ. Int.*, vol. 59, pp. 244–254, 2013.
[10] B. Aydin, V. Akkineni, and R. Angryk, "Mining spatiotemporal co-occurrence patterns in non-relational databases," *GeoInformatica*, vol. 20, no. 4, pp. 801–828, 2016.
[11] 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 Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 559–568, Jan. 2016.
[12] R. J. Bennett. *Spatial Time Series*. London, United Kingdom: Pion, 1979.
[13] J. Y. Blaise, and I. Dudek, "Spotting temporal co-occurrence patterns: the historySkyline visual metaphor," in *Proc. 1st Int. Conf. Metrology Archaeology*, 2015, pp. 378–383.
[14] B. C. Cappers, and J. J. van Wijk, "Exploring multivariate event sequences using rules, aggregations, and selections," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 532–541, Jan. 2018.
[15] N. C. Castro, and P. J. Azevedo, "Significant motifs in time series," *Statistical Anal. Data Mining*, vol. 5, no. 1, pp. 35–53, 2012.
[16] M. Celik, S. Shekhar, J. P. Rogers, and J. A. Shine, "Mixed-drove spatiotemporal co-occurrence pattern mining," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 10, pp. 1322–1335, Oct. 2008.
[17] M. Celik, N. Azginoglu, and R. Terzi, "Mining periodic spatiotemporal co-occurrence patterns: A summary of results," in *Proc. Int. Symp. Innovations Intell. Syst. Appl.*, 2012, pp. 1–5.
[18] Q. Chen, Y. Chen, D. Liu, C. Shi, Y. Wu, and H. Qu., "Peakvizor: Visual analytics of peaks in video clickstreams from massive open online courses," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 10, pp. 2315–2330, Oct. 2016.
[19] S. Chen, X. Yuan, Z. Wang, C. Guo, J. Liang, Z. Wang, and J. Zhang, "Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, 1, pp. 270–279, Jan. 2016.
[20] DECISIVe 2017, "Dealing with cognitive biases in visualisations: A VIS 2017 workshop," [Online]. Available: http://decisive-workshop.dbvis.de/
[21] F. Du, C. Plaisant, N. Spring, and B. Shneiderman, "EventAction: Visual analytics for temporal event sequence recommendation," in *Proc. IEEE Conf. Vis. Anal. Sci. Technol.*, 2016, pp. 61–70.
[22] S. Elzen, D. Holten, J. Blaas, and J. J. Wijk, "Reducing snapshots to points: A visual analytics approach to dynamic network exploration," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 1–10, Jan. 2016.
[23] J. H. Faghmous, and V. Kumar, "Spatio-temporal data mining for climate data: Advances, challenges, and opportunities," *Data Mining and Knowledge Discovery for Big Data*. Berlin, Germany: Springer, 2014, pp. 83–116.
[24] M. Gregory, and B. Shneiderman, "Shape identification in temporal data sets," in: J. Dill, R. Earnshaw, D. Kasik, J. Vince, P. C. Wong. *Expanding the Frontiers of Visual Analytics and Visualization*, Berlin, Germany: Springer, 2012, pp. 305–321.
[25] A. Greiner, W. Semmler, and G. Gong, "The forces of economic growth: A time series perspective," *Princeton University Press*, 2016.
[26] V. Guralnik, and J. Srivastava, "Event detection from time series data," in *Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 1999, pp. 33–42.
[27] J. Hofmann, M. Größler, M. Rubio-Sánchez, P. P. Pichler, and D. J. Lehmann, "Visual exploration of global trade networks with time-dependent and weighted hierarchical edge bundles on GPU," *Comput. Graph. Forum*, vol. 36, no. 3, pp. 273–282, 2017.
[28] H. Izakian, and W. Pedrycz, "Anomaly detection and characterization in spatial time series data: A cluster-centric approach," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 6, pp. 1612–1624, Dec. 2014.
[29] P. C. Kyriakidi, N. L. Miller, and J. Kim, "A spatial time series framework for modeling daily precipitation at regional scales," *J. Hydrology*, vol. 297, no. 1–4, pp. 236–255, 2004.
[30] T. von Landesberger, S. Bremm, N. Andrienko, G. Andrienko, and M. Tekusova, "Visual analytics methods for categoric spatiotemporal data," in *Proc. IEEE Conf. Visual Anal. Sci. Technol.*, 2012, pp. 183–192.
[31] T. von Landesberger, F. Brodkorb, P. Roskosch, N. Andrienko, G. Andrienko, and A. Kerren, "MobilityGraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, 1, pp. 11–20, Jan. 2016.
[32] J. Li, Z. P. Meng, M. L. Huang, and K. Zhang, "An interactive visualization approach to the overview of geoscience data," *J. Vis.*, vol. 20, no. 3, pp. 433–451, 2017.
[33] D. B. Neill, "Expectation-based scan statistics for monitoring spatial time series data," *Int. J. Forecasting*, vol. 25, no. 3, pp. 498–517, 2009.
[34] J. Li, Z. Xiao, H. Q. Zhao, Z. P. Meng, and K. Zhang, "Visual analytics of smogs in China," *J. Vis.*, vol. 19, no. 3, pp. 461–474, 2016.
[35] L. Li, X. Su, Y. Zhang, Y. Lin, and Z. Li., "Trend modeling for traffic time series analysis: An integrated study," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 3430–3439, Dec. 2015.
[36] J. Li, K. Zhang, and Z. P. Meng, "Vismate: Interactive visual analysis of station-based observation data on climate changes," In *Proc. IEEE Conf. Visual Anal. Sci. Technol.*, 2014, pp. 133–142.
[37] S. Lohmann, M. Burch, H. Schmauder, and D. Weiskopf, "Visual analysis of microblog content using time-varying co-occurrence highlighting in tag clouds," in *Proc. Int. Working Conf. Adv. Visual Interfaces*, 2012, pp. 753–756.
[38] X. Ma, Z. Liu, X. Zhao, L. Tian, and T. Wang, "The spatial and temporal variation of Haze and its relativity in Beijing-Tianjin-Hebei Region," *Areal Res. Develop.*, vol. 35, no. 2, pp. 134–138, 2016.
[39] Ministry of environmental protection of the People's Republic of China. Ambient Air Quality Standards (on trial). *National Environmental Protection Standards of the People's Republic of China* (GB3095-2012), 2012.

[40] M. Monroe, R. Lan, C. Plaisant, and B. Shneiderman, "Temporal event sequence simplification," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2227–2236, Dec. 2013
[41] D. J. Peuquet, A. C. Robinson, S. Stehle, F. A. Hardisty, and W. Luo, "A method for discovery and analysis of temporal patterns in complex event data," *Int. J. Geographical Inf. Sci.*, vol. 29, no. 9, pp. 1588–1611, 2015.
[42] F. Qian, Q. He, and J. He, "Mining spread patterns of spatiotemporal co-occurrences over zones," in *Proc. Int. Conf. Computat. Sci. Its Appl.*, 2009, pp. 677–692.
[43] H., Qu, W. Y. Chan, A. Xu, K. L. Chung, and P. Guo, "Visual analysis of the air pollution problem in Hong Kong," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, 6, pp. 1408–1415, Nov./Dec. 2007.
[44] J. C. Roberts, "State of the art: Coordinated & multiple views in exploratory visualization," in *Proc. 5th Int. Conf. Coordinated Multiple Views Exploratory Vis.*, 2007, pp. 61–71.
[45] H. Rosling. Gapminder: World. [Online]. Available: http://www.gapminder.org/world, 2008.
[46] S. Shekhar, and Y. Huang, "Discovering spatial co-location patterns: A summary of results," in *Proc. Int. Symp. Spatial Temporal Databases*, 2001, pp. 236–256.
[47] G. Shurkhovetskyy, N. Andrienko, G. Andrienko, and G. Fuchs, "Data abstraction for visualizing large time series," *Comput. Graph. Forum*, vol. 37, pp. 125—-144, 2018.
[48] D. G. Streets, J. S. Fu, C. J. Jang, J. Hao, K. He, X. Tang, Y. Zhang, Z. Wang, Z. Li, Q. Zhang, L. Wang, B. Wang, and C. Yue, "Air quality during the 2008 Beijing Olympic Games," *Atmospheric environment*, vol. 41, no. 3, pp. 480–492, 2007.
[49] C. A. Thayer, "The Structure of Vietnam-China Relations, 1991-2008," in *Proc. Int. Conf. Vietnamese Studies*, 2008, pp. 4–7.
[50] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko, "Stacking-based visualization of trajectory attribute data," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2565–2574, Dec. 2012.
[51] A. Unger, N. Dräger, M. Sips, and D. J. Lehmann, "Understanding a sequence of sequences: Visual exploration of categorical states in lake sediment cores," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 66–76, Jan. 2018.
[52] A. C. Valdez, M. Ziefle, and M. Sedlmair, "A Framework for studying biases in visualization research," in *Proc. Workshop Dealing Cognitive Biases Vis.*, 2017, [Online]. Available: http://decisive-workshop.dbvis.de/wp-content/uploads/2017/09/0118-paper.pdf
[53] M. Waldner, W. Puff, A. Lex, M. Streit, and D. Schmalstieg, "Visual links across applications," in *Proc. Graph. Interface*, 2010, pp. 129–136.
[54] E. Wall, L. Blaha, L. Franklin, and A. Endert, "Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics," *IEEE Visual Anal. Sci. Technol.*, 2017.
[55] W. Wu, J. Xu, H. Zeng, Y. Zheng, H. Qu, B. Ni, M. Yuan, and L. M. Ni., "Telcovis: Visual exploration of co-occurrence in urban human mobility based on telco data," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 935–944, Jan. 2016.
[56] D. Yankov, E. Keogh, J. Medina, B. Chiu, and V. Zordan, "Detecting time series motifs under uniform scaling," in *Proc. Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2007, pp. 844–853
[57] J. S. Yoo, S. Shekhar, J. Smith, and J. P. Kumquat, "A partial join approach for mining co-location patterns," in *Proc. ACM Int. Workshop Geographic Inf. Syst.*, 2004, pp. 241–249.
[58] J. S. Yoo, and S. Shekhar, "A joinless approach for mining spatial colocation patterns," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1323–1337, Nov. 2006.
[59] P. Zhang, Y. Huang, S. Shekhar, and V. Kumar, "Correlation analysis of spatial time series datasets: A filter-and-refine approach," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*, 2003, pp. 563–563.
[60] Z. Zhang, and W. Wu, "Composite spatio-temporal co-occurrence pattern mining," in *Proc. Wireless Algorithms, Syst. Appl.*, 2008, pp. 454–465.

Jie Li is an assistant professor of the school of computer software with Tianjin University. His current research interests include visualization and visual analytics, in particular, for environmental sciences, public security, and social media. He has served as the program chair of *VINCI 2017* and published more than 20 papers in visualization conferences and journals, including *IEEE VIS*, *Journal of Computer*, etc.

Siming Chen received the PhD degree from Peking University. He is a research scientist with Fraunhofer Institute IAIS and a postdoc researcher of the University of Bonn in Germany. His research interests include visual analytics of social media, cyber security and spatial temporal data. He published several papers in IEEE VIS, IEEE TVCG, EuroVis, etc. More information can be found in http://simingchen.me.

Kang Zhang is a professor and director of Visual Computing Lab, the Department of Computer Science with the University of Texas, Dallas. He has authored and edited seven books, and published more than 80 journal papers in the areas of computational aesthetics, information visualization, and visual languages. He is on the editorial boards of several reputable journals. His home page is at utd.edu/∼kzhang.

Gennady Andrienko is a lead scientist focusing on visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems in Germany and part-time professor with City University London. He is an associate editor of two journals, *Information Visualization* and *International Journal of Cartography*., and editorial board member of *Cartography and Geographic Information Science* and *Cartographica*.

Natalia Andrienko is a lead scientist focuising on visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems in Germany and part-time professor with City University London. Results of her research on spatio-temporal data analysis have been published in two monographs "*Exploratory Analysis of Spatial and Temporal Data: a Systematic Approach*" (Springer 2006) and "*Visual Analytics of Movement*" (Springer 2013). She is associate editor of *IEEE Transactions on Visualization and Computer Graphics*.