DOI: 10.1111/cgf.14535
Eurographics Conference on Visualization (EuroVis) 2022
R. Borgo, G. E. Marai, and T. Schreck
(Guest Editors)

Volume 41 (2022), Number 3

AirLens: Multi-Level Visual Exploration of Air Quality Evolution in Urban Agglomerations

Dezhan Qu[1,2], Cheng Lv[1,2], Yiming Lin[1,2], Huijie Zhang (Corresponding author)[1,2†] and Rong Wang[1,2]

[1]Northeast Normal University, School of Information Science and Technology, Changchun, China
[2]Key Laboratory of Intelligent Information Processing of Jilin Universities, Changchun, China

Abstract

The precise prevention and control of air pollution is a great challenge faced by environmental experts in recent years. Understanding the air quality evolution in the urban agglomeration is important for coordinated control of air pollution. However, the complex pollutant interactions between different cities lead to the collaborative evolution of air quality. The existing statistical and machine learning methods cannot well support the comprehensive analysis of the dynamic air quality evolution. In this study, we propose AirLens, an interactive visual analytics system that can help domain experts explore and understand the air quality evolution in the urban agglomeration from multiple levels and multiple aspects. To facilitate the cognition of the complex multivariate spatiotemporal data, we first propose a multi-run clustering strategy with a novel glyph design for summarizing and understanding the typical pollutant patterns effectively. On this basis, the system supports the multi-level exploration of air quality evolution, namely, the overall level, stage level and detail level. Frequent pattern mining, city community extraction and useful filters are integrated into the system for discovering significant information comprehensively. The case study and positive feedback from domain experts demonstrate the effectiveness and usability of AirLens.

CCS Concepts

• ***Human-centered computing*** → ***Visual analytics; Geographic visualization;*** *Information visualization;*

1 Introduction

Air pollution has become a major environmental issue around the world. In the report published by WHO in 2019 ‡, air pollution was listed as the greatest environmental risk to human health. Since air quality generally evolves within a specific region collaboratively, urban agglomerations are divided for air pollution control, according to the geographical and economic conditions. Understanding the complex air quality evolution in the urban agglomeration is one of the important prerequisites for controlling the air pollution jointly. However, the air quality evolution in an urban agglomeration is a complex spatiotemporal process, which is difficult to be analyzed.

Currently, existing works of spatiotemporal analysis for air quality can be classified into statistical, learning-based and visualization-based methods. Statistical methods generally compute metrics for various goals, such as spatial autocorrelation [SZF*19], significance test and causality test [LCCC17]. Learning-based methods can reveal more complex information, such as spatial clusters, temporal patterns [MGDVT18] and spatiotemporal association rules [LC06]. However, these methods focus on static metrics or patterns, and are weak in analyzing the dynamic and complex evolution of air quality in urban agglomeration. This motivates us to adopt visualization-based methods, which allow experts to intuitively inspect the complex spatiotemporal characteristics of air quality data. Previous visualization studies for air quality data mainly focused on analyzing variable correlations [GTC*19], spatial clusters [ZYL*17] and pollution propagation [DWC*19]. Some works [QCX*07, LXZ*16] also provided overview visualizations from multiple perspectives. However, to the best of our knowledge, rare works in the visualization community have analyzed the air quality evolution in urban agglomeration, which is an immediate problem required to be examined. Developing an effective visual analytics system for air quality evolution in urban agglomeration faces the following challenges:

Effective analysis of air quality evolution. Analyzing air quality evolution in the urban agglomeration requires to incorporate multiple key topics concerned by experts, such as multivariate patterns, city relationships and evolution patterns. Proposing an effective framework that supports comprehensive analysis is challenging.

Visualization of large-volume spatiotemporal data. Displaying the evolution of a dataset with multivariate and spatiotemporal information will cause scalability problems, especially over a long time period. How to encode such complex information and ensure the effectiveness and scalability is a great challenge.

† Corresponding author: Huijie Zhang (zhanghj167@nenu.edu.cn).
‡ www.paho.org/en/news/17-1-2019-ten-threats-global-health-2019

Flexible and efficient exploration of complex patterns. Experts desiderate to explore and understand the data-driven analysis results, thus they can discover new knowledge and verify hypotheses. Designing a system with flexible, efficient and reasonable interactions for the complex dataset is an essential yet challenging task.

To address the first challenge, we extract the analysis tasks with domain experts and propose an analysis framework that contains the *pollutant pattern inspection module* and the *evolution analysis module*, which allows experts to explore the data from multiple aspects. Moreover, the proposed framework supports multi-level analysis, which addresses the second challenge. It can summarize the complex data from pollutant, spatial and temporal perspectives, and the details can also be explored. For the third challenge, we solve it by developing *AirLens*, an interactive system containing multiple coordinated views and novel designs. It allows domain experts to adjust the parameters of automatic methods and visualizations flexibly, and provides visual feedback immediately. Through the system, experts can discover insightful patterns and verify their hypotheses. Our contributions are refined as follows:

- We characterized the problem domain of analyzing air quality evolution in urban agglomeration during a long period, and summarized a set of analysis tasks with domain experts.
- We proposed a multi-level framework that reveals complex collaborative evolution of air quality effectively and supports comprehensive analysis of significant patterns from multiple aspects.
- We proposed *AirLens*, a flexible visual analytics system for environmental experts to understand and explore complex data patterns, through novel visualizations and rich interactions.

2 Related Work

2.1 Air quality analysis and visualization

Air quality analysis is an important topic for various domains. From the perspective of used methods, it can be classified into model-based methods and data-driven methods. For the model-based methods, air quality numerical models were developed based on atmospheric physical and chemical processes for different applications. The common used models include CMAQ [BS06], WRF-Chem [GPS*05] and HYSPLIT [SDR*15]. Environmental experts could conduct simulations using these models and verify their hypotheses. However, uncertainty inevitably existed in these models, and the simulation results were largely affected by the settings of initial conditions and parameters. The huge time cost of the models also hindered experts from conducting comprehensive analysis.

With the extensive monitoring of air quality in recent years, researchers had a good opportunity to study air pollution from the data-driven perspective. Methods based on various theories have been proposed. Shen et al. [SZF*19] utilized statistical methods to quantify the spatiotemporal characteristics of air quality data. Complex network theories [FWX*16, QDL*21] were introduced to analyze the topological features of air quality data and divide the temporal data into flexible periods. Urban computing studies [ZCWY14, ZYL*15] integrated multi-source data, such as weather and traffic, to analyze and predict air quality. Recently, deep learning methods [QLKL19, ZDZZ21, HLZ*21] have been extensively used in air quality prediction. Compared with the model-based methods, these approaches were commonly more efficient. Meanwhile, their effectiveness and accuracy have also been proved. However, most of these methods focused on static analysis tasks and lacked the interactive analysis process. Thus, they were weak in analyzing the complex and dynamic air quality evolution. Moreover, they generally could not consider domain knowledge well, which made the results hard to be understood by domain experts.

Visualization is an effective way to solve the limitations of fully automatic methods, which exploits human's efficient visual perception to seamlessly bring users into the analysis process. Users can iteratively refine the data and explore the hidden information, which can facilitate users' understanding of complex data [TC05]. Based on air quality data, visual analytics systems have been proposed to solve various spatiotemporal problems in the environmental field, such as the evolution of spatial clusters [ZYL*17, GYL*17], propagation patterns of air pollution [DWC*19, RWZ*20], temporal evolution patterns [QLR*20, ZRL*19], anomaly detection and examination [LVG*21], and co-occurrence analysis [LCZ*19]. Nevertheless, few works have focused on analyzing the collaborative evolution of air quality in urban agglomeration, due to the complex data structure. In this paper, we design a comprehensive visual analytics system with multi-level visualizations to explore the complex air quality evolution in urban agglomeration.

2.2 Visualization of spatiotemporal data and event sequence

Air quality data has spatial, temporal and multivariate characteristics, which can be considered as multivariate spatial time-series. The techniques for time-series visualization, such as the spiral-shaped visualization [WAM01], temporal stacked charts [WWS*16] and time-embedded parallel coordinates [GRPF16], can also be used to analyze air quality data. However, these methods cannot encode all the information in one view. Another dimensional information needed to be displayed by other views. Some methods attempted to present spatial, temporal and attribute information through one visualization, such as global radial map [LZM14], glyphs on maps [AAFW17], methods based on space-time cube [Kra08, LWSY20], and small multiples [BDM*17].

To reduce the data complexity and facilitate experts' understanding, we summarize the multiple pollutants of air quality data into several pollutant patterns. Thus, the temporal air quality data can be regarded as event sequences. The recent survey [GGJ*21] has summarized the visual analysis methods of event sequences comprehensively. For sequence analysis, frequent pattern mining and sequence summarization are two kinds of widely used methods to discover common patterns of event sequences. VMSP [FVWGT14], SPAM [AFGY02] and MDL-based algorithms [CXR18] have been used in the visual analytics systems [CYP*20, PW14, WGW*20] for various objectives. Guo et al. [GXZ*18, GJG*19] proposed two stage-based summarization methods for sequence visualization. CoreFlow [LKD*17] could automatically extract the branching patterns in event sequences and summarize the event sequences as a tree structure. For visualization of event sequences, flow-based visualization is an effective visual representation of large-scale sequence data, with the Sankey-based structure [WLS*21, WLG*21] or tree-based structure [LKD*17]. Matrix-based and list-based visualizations were also used to provide a scalable overview of event sequences [ZLD*15] and compare different event sequences

[WGW*20]. In this paper, using multiple techniques of showing event sequences, we design multi-level visualizations with geographical information to display the air quality evolution in the urban agglomeration.

3 Overview

3.1 Data description and pre-processing

The air quality dataset is collected from the data platform of China National Environmental Monitoring Centre, covering more than 1400 observation stations in 367 cities, from January 5, 2015 to January 3, 2016. The data can reflect the air quality under different weather conditions throughout the year. For each station, the data are hourly records of the concentration of six pollutants containing $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO and O_3.

Since many missing values exist in the data and our goal is analyzing the air quality evolution during a long period and within a large region, the experts set the analysis granularity as city and day. Thus, for each pollutant, we computed its daily average of the hourly concentrations among all the stations in each city, which is a standard method in environmental field. Besides, the concentrations of different pollutants have different units, which is difficult to identify and compare their pollution levels. Thus, we calculated the individual air quality index (IAQI) [EPD16] of different pollutants, thus making all the pollutants have the same range [0, 500]. Moreover, for each day in each city, the air quality index (AQI) is the maximal IAQI among the six pollutants. Based on the AQI, air quality can be divided into six levels.

3.2 Design process and task analysis

In the past year, we worked closely with two environmental experts. E_A is a scholar of environmental evaluation, who is interested in the coordinated control of air pollution in urban agglomeration. E_B is an atmospheric physics expert from the same institute as E_A. Guided by the nine-stage design study methodology framework [SMM12], we first conducted literature reviews about the analysis and visualization of spatiotemporal air quality data. Then, we had frequent discussions with domain experts through video meetings and emails to iteratively refine the analysis tasks.

In the first discussion, E_A and E_B introduced the study pipeline of their domains and put forward their high-level requirements. E_A wanted to understand the spatiotemporal characteristics of air quality in the five key urban agglomerations in China, to write air quality reports. E_B wished to find some significant patterns for further study. We also described how visual analytics can help them from the data-driven perspective. In the next two months, through designing multiple overall views and iteratively discussing with the experts, we obtained some initial analysis tasks. During the system development, the prototype was developed and improved iteratively based on the experts' feedback. The analysis tasks were also refined and finally summarized into two modules, namely, *pollutant pattern inspection module* and *evolution analysis module*.

The tasks of the *pollutant pattern inspection module* support experts in summarizing and exploring the typical pollutant patterns and their spatiotemporal characteristics.

T1. What are the typical pollutant patterns among the data? The multiple air pollutants have complex associations with each other, which are influenced by topography and weather conditions. Experts wish to obtain the typical pollutant patterns in the specific urban agglomeration during a period, which helps them verify and infer the latent pollutant associations.

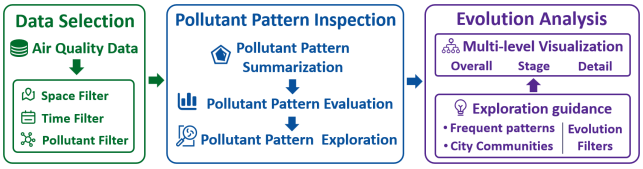

Figure 1: *System overview. AirLens has three modules: data selection, pollutant pattern inspection and evolution analysis.*

T2. What are the spatiotemporal characteristics of the typical pollutant patterns? Experts want to further explore the spatiotemporal characteristics of these patterns, to understand the pollutant patterns deeply.

The tasks of *evolution analysis module* support multi-level and multi-facet visual exploration of the air quality evolution.

T3. What are the latent stages during the whole period? Due to the large volume and complexity of air quality data, it is difficult for experts to obtain detailed information from an overview analysis. Dividing the period into stages adaptively and revealing the key information can help experts understand the evolution effectively with a medium level of details and acceptable data volume.

T4. What are the city communities within an urban agglomeration? The cities in an urban agglomeration can be divided into small communities according to their latent associations. Experts want to discover the latent city communities in a data-driven way and explore the community-level air quality evolution.

T5. What are the frequent patterns of air quality evolution within a stage? During the exploration of evolution trends, experts also wish to know if there are any patterns that frequently occur inside a stage, which can facilitate the perception and understanding of the evolution trends.

T6. What are the evolution details inside a stage? The air quality evolution is related to complex information regarding geographical locations, time and pollutants. To inspect and understand the air quality evolution effectively, experts require a view that presents the evolution details inside a stage with comprehensive context.

3.3 System overview

Figure 1 presents the workflow of *AirLens*. It is a web application that comprises three modules, namely, data selection, pollutant pattern inspection and evolution analysis. The *data selection module* allows experts to select the target urban agglomeration, time range and pollutants. For the selected data, the *pollutant pattern inspection module* online summarizes the pollutant patterns (T1), computes their evaluation metrics (T1) and shows their spatiotemporal distributions (T2). Based on the air quality evolution represented by the pollutant patterns, the *evolution analysis module* divides the whole period into stages adaptively (T3) and thus provides a multi-level visualization of the evolution trends (T3, T6). City communities (T4), frequent patterns (T5) and multiple evolution filters are provided to help experts explore the complex evolution trends.

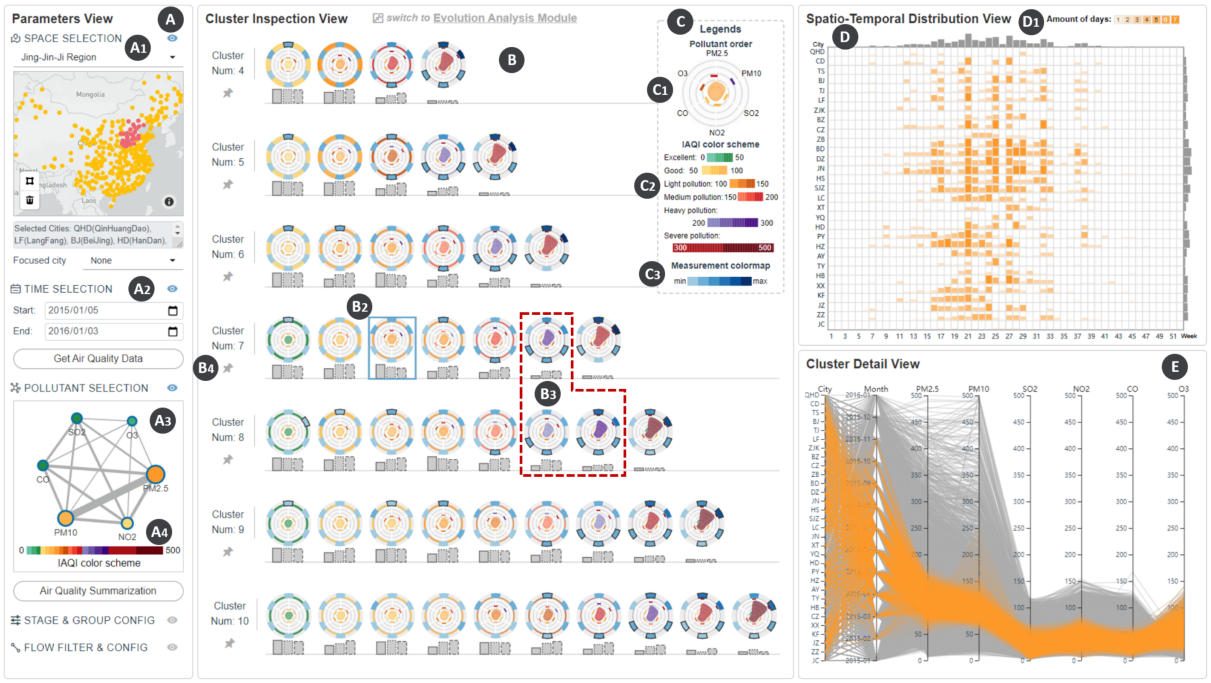

Figure 2: *Visual design of the pollutant pattern inspection module. (A) Parameters view for selecting data. (B) Cluster inspection view for checking all the clustering results. (C) Legends for AirLens. (D) Spatiotemporal distribution view shows spatiotemporal distribution of a focused cluster. (E) Cluster detail view is a PCP that shows the detailed data items of a focused cluster.*

The architecture of *AirLens* consists of a back-end supported by Flask and MongoDB, and a front-end implemented by Vue.js and D3.js. The back-end supports the computations of data processing and all the algorithms used in the *AirLens*. The front-end interface implements the data filters, visualizations and interactions.

4 System design

4.1 Data selection module

The *data selection module* (Figure 2-A) helps experts select the data of interest from three aspects: space, time and pollutants.

The *space selector* (Figure 2-A1) provides two ways to select the target urban agglomeration. One way is to select a specified urban agglomeration from the drop-down menu. The other way is to select a set of cities by drawing a polygon on the map. The cities selected in both ways are highlighted on the map. The *time selector* (Figure 2-A2) supports the setting of the start date and end date.

The *pollutant selector* (Figure 2-A3) is a force-directed node-link diagram that reveals the pollutant significance and associations in the selected region and period, which helps experts to select the pollutants of interest. Each node denotes a pollutant, and the color encodes the mean of the pollutant's IAQIs. For meeting the experts' habit of perceiving pollution levels, we design the *IAQI color scheme* according to the technical regulation [EPD16] that specifies the standard colors for six pollution levels. To encode the IAQI in detail, we extended each standard color to a sequential colormap as shown in Figure 2-A4 and C2, which were selected from ColorBrewer using the colorblind-safe mode. This color scheme is used to encode the IAQI in all the views of *AirLens*.

Moreover, entropy [Sha48] and mutual information [CT90] are computed to reveal the information content of each pollutant's IAQIs and the associations between each pair of pollutants, as encoded by the node size and link width, respectively. Based on these metrics, experts can select the pollutants of interest by clicking the nodes. The selected nodes are highlighted with a blue border, while the deselected nodes are with gray border. All the pollutants are selected by default, while the deselected pollutants are not computed and not shown in the system.

Justification: Initially, we used MDS [BG05] to project the pollutant nodes into 2D space based on the similarities between their IAQI distributions, as shown in Figure 4-A. However, the positions of projected nodes may hinder the perception of edges. Thus, we adopted the force-directed layout that could better display edges.

4.2 Pollutant pattern inspection module

After selecting the target urban agglomeration and period, the *pollutant pattern inspection module* (Figure 2) is used to summarize the air quality data into typical pollutant patterns (T1), evaluate their spatiotemporal characteristics (T2), and help experts inspect the pollutant patterns through visual exploration.

Pollutant pattern summarization. To facilitate the understanding of multiple pollutants, we summarize the air quality data as several representative pollutant patterns by clustering, which is an effective way for multivariate data summarization. Multiple methods have been considered, including K-Means [Mac67], DBSCAN [EKSX96] and hierarchical clustering [Fox91]. We choose K-Means, since it is intuitive for experts and very efficient. To help

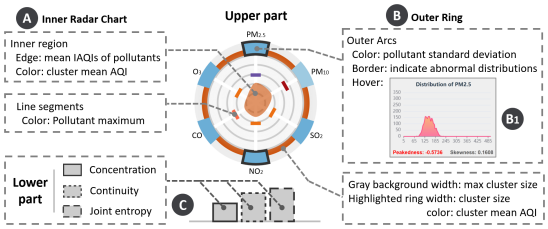

Figure 3: *Design of cluster glyph that has an upper part comprised of inner radar chart (A) and outer ring (B), and a lower part (C).*

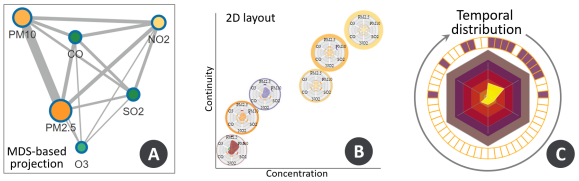

Figure 4: *Alternatives for (A) layout of pollutant selector, (B) layout of cluster inspection view and (C) design of cluster glyph.*

experts obtain the satisfied clustering result and reduce the burdens of testing different cluster numbers, we propose a multi-run clustering strategy that juxtaposes the results of K-Means using different cluster numbers, as the rows in the *cluster inspection view* (Figure 2-B). In each row, the clusters, namely the pollutant patterns, are arranged sequentially, and the key information of each cluster is revealed by the designed glyph. Through multiple experiments and discussions with experts, the cluster numbers are set from 4 to 10, which can handle various situations, including the regions and periods with very stable and unstable air quality.

Pollutant pattern evaluation and glyph design. We compute multiple metrics to describe and evaluate the clusters. The descriptive metrics reveal the main information of each pollutant, including mean, maximum, standard deviation, as well as peakedness and skewness of the IAQI distribution. Furthermore, three metrics are introduced to evaluate the significance of a cluster's spatiotemporal distribution, namely concentration, continuity [WSH13] and joint entropy. Concentration and continuity describe the intensive degree and connection degree of cluster occurrence in the spatial and temporal distribution, respectively. Joint entropy evaluates the amount of information from the spatiotemporal distribution of the cluster.

To present these metrics intuitively, we design the cluster glyph as shown in Figure 3. Since the descriptive and the evaluative metrics represent different types of information, we divide the cluster glyph into an upper part and a lower part. Considering the multiple pollutants, the upper part (Figure 3-A and B) is designed as a radar chart with an outer ring to show the descriptive metrics. In the inner radar chart, the area displays the mean IAQIs of different pollutants. Furthermore, the line segments display the maximal IAQIs of pollutants. To facilitate the perception of pollution levels, the area and line segments are colored with the *IAQI color scheme*. The outer ring comprises a gray background, a highlighted ring and the arcs corresponding with the pollutants. The width of the highlighted ring represents the cluster size, and its color is the same as the inner area. The width of the gray background indicates the maximal cluster size among all the clusters computed using different cluster numbers. The arcs show the standard deviation of pollutants, using the *measurement colormap* as shown in Figure 2-C3. Besides, some arcs are with black borders, which indicates abnormal peakedness and skewness. Experts can hover on the arc to check the distribution of the pollutant's IAQIs and the values of peakedness and skewness, as shown in Figure 3-B. Since the evaluative metrics are all quantitative, they are encoded by three bars with different border styles in the lower part (Figure 3-C).

Context views. We develop a spatiotemporal distribution view (Figure 2-D) and a parallel coordinates (PCP) view (Figure 2-E), to help experts examine the spatiotemporal distribution and data details of a focused cluster, by clicking the corresponding cluster glyph. For the spatiotemporal distribution view, the x-axis represents the weeks, and the y-axis represents the cities in the urban agglomeration. These cities are sorted by the 1D projected coordinates of their longitudes and latitudes using MDS, which can reflect the city adjacencies and better interpret the continuity of the spatiotemporal distribution. The bar's height and opacity both encode the number of days that the pollution pattern occurs in this week at the city, as shown in the legend in Figure 2-D1. The PCP comprises the axes of the city, month and pollutants. All the data items are shown as the background with gray lines. When a cluster is selected, the data items of this cluster will be highlighted using the color of the cluster's mean AQI. The PCP also supports filtering and reordering of the axes, which helps users explore the whole data. After exploring and comparing different groups of clusters, users can select a satisfied cluster group as the pollutant patterns, by clicking the icon as shown in Figure 2-B4 and further analyze the evolution among these pollutant patterns in the *Evolution analysis module*.

Justification: For the layout of cluster glyphs, we first show them in a 2D space with the axes of concentration and continuity, as shown in Figure 4-B. However, the 2D layout does not well adapt to different cluster numbers. When the cluster number is small, the view is space ineffective. When the cluster number is large, it will result in visual clutter. Thus, we use the 1D layout to show the cluster glyphs sequentially and show the concentration and continuity by the bars in the lower part of the glyph, which are more scalable and the remaining space can be used to display the legends. For the glyph design, we also try to display the temporal distribution of the cluster by the arcs shown in Figure 4-C. However, the spatial information can not be encoded. Thus, we display the cluster's spatiotemporal distribution by a separate view, and replace the arcs with the outer ring that shows other useful information.

4.3 Evolution analysis module

After the pollutant patterns are selected, the original multivariate spatiotemporal data can be simplified into a spatial event sequence with uniform time steps, where each pollutant pattern is regarded as an event. The *evolution analysis module* (Figure 5) helps experts understand the evolution in air quality event sequences, by supporting the multi-level (T3, T6) exploration and the guidance based on city communities (T4) and frequent patterns (T5).

4.3.1 Overall-level view

To provide an overview of the whole period, we use a temporal stacked bar chart to show the distribution of pollutant patterns with-

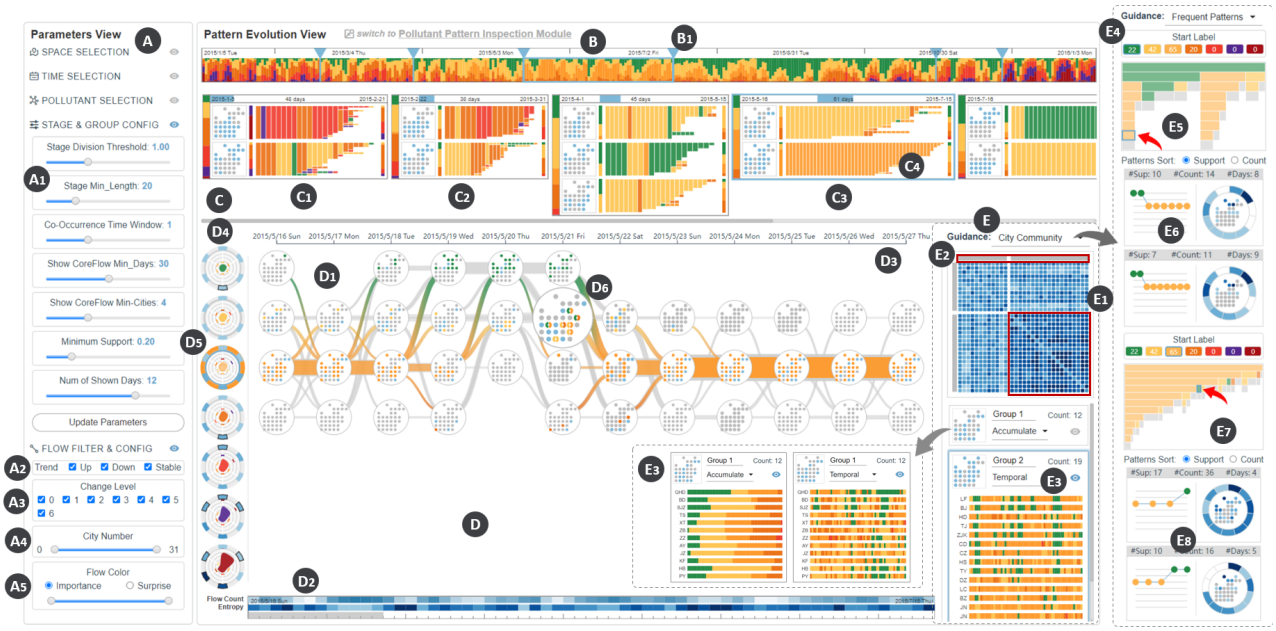

Figure 5: *Visual design of evolution analysis module. (A) Parameters view for setting algorithm parameters and flow filters. (B) Overall-level view is a temporal stacked bar chart that shows pollutant pattern distribution among the urban agglomeration for each day. (C) Stage-level view contains some stage cards that show major trends for city communities. (D) Spatiotemporal flow chart shows detailed evolution trends in the focused stage with a spatiotemporal context. (E) Guidance part based on city community and frequent patterns.*

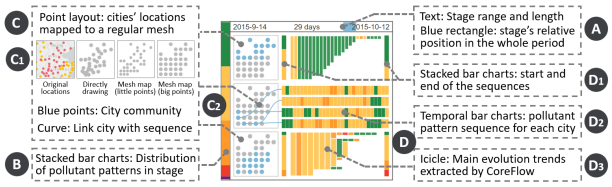

Figure 6: *Design of stage card. (A) Stage range bar. (B) Pollutant pattern distribution in the stage. (C) City location components show city communities. (D) Main trend component with two modes.*

in the urban agglomeration for each day. As shown in Figure 5-B, the x-axis represents days. The height of a bar encodes the number of cities where a pollutant pattern occurs on the day, and the bar color is the same as the corresponding cluster glyph's inner area.

4.3.2 Stage-level view

To help experts understand the air quality evolution with more details, we design the *stage-level view* (Figure 5-C) that divides the whole data into meaningful stages (**T3**) and presents the main evolution trends for latent city communities (**T4**).

Computation. From the temporal perspective, we divide the whole period into latent stages based on an unsupervised stage extraction algorithm [GXZ*18] for event sequences. It encodes events through word embedding and extracts stages based on the semantic meaning of event sequences. To further refine the evolution from the spatial perspective, we construct the city relationships among the urban agglomeration based on co-occurrence and successive occurrence of pollutant patterns in a time window. On this basis, we divide the urban agglomeration into city communities using the efficient Louvain algorithm [BGLL08], which detects communities by maximizing a modularity score for each community.

Visual design. To reveal the major evolution trends in a stage, we design the *stage card* as shown in Figure 6. It presents the general stage information including the stage range (Figure 6-A) and distribution of pollutant patterns (Figure 6-B), as well as the evolution trends of each city community. Specifically, the *city location component* (Figure 6-C) displays the relative locations of the cities in each city community. To avoid the occlusion of city points, we map them to their most closed locations on a regular square mesh, which can largely maintain the proximity between cities (Figure 6-C1). The *main trend component* (Figure 6-D) provides two modes of visual representations. If the stage length is not greater than 30 and the size of the city community is less than 5, we directly show the pollutant pattern sequence of each city (Figure 6-D2) and link these sequences to corresponding cities by curves (Figure 6-C2). Otherwise, to avoid visual clutter, we introduce the CoreFlow algorithm [LKD*17] to extract main evolution trends, which can provide a high-level summary of long sequences by generating a tree structure using key events as nodes. Its result can be shown by an icicle plot (Figure 6-D3), where rectangular partitions represent key pollutant patterns in the sequence, and the partition's height indicates the number of cities containing this pollutant pattern. The hierarchy indicates the evolution order in the sequence. To highlight the important information at the start and end of the stage, we add stacked bar charts (Figure 6-D1) to represent their pollutant pattern distributions on the left and right sides of the icicle plot. The *stage-level view* (Figure 5-C) arranges the *stage cards* in temporal order, where experts can click the stage of interest to inspect details. Stage marks (Figure 5-B1) are also added onto the *overall-level view*.

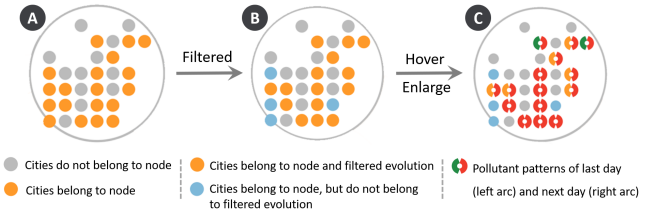

Figure 7: *Visual design of spatial node. (A) Normal mode of spatial node. (B) Spatial node after filtering by city communities, frequent patterns and flow filters. (C) Spatial node after hovering.*

Justification: During designing the *stage card*, we have compared our final design with several alternatives of showing evolution trends, including: (1) directly showing original sequences with colored bars, (2) showing original sequences with highlighting the start and end of sequences, and (3) only showing the icicle plot of CoreFlow. Among them, the designs that show original sequences are hard to perceive for long stages. Besides, the design that only shows the icicle plot may lose the information of stage start and end. Thus, we embed the icicle plot between the stacked bar charts that summarize the start and end of the sequences.

4.3.3 Detail-level view

Experts can inspect the detailed evolution among a stage of interest in the *detail-level view* (Figure 5-D and E), which comprises the *spatiotemporal exploration part* and the *guidance part*.

The **Spatiotemporal exploration part** allows experts to explore the detailed evolution in a stage with a spatiotemporal context. Due to the contradiction between limited screen space and a large amount of details, we design a local *spatiotemporal flow chart* (Figure 5-D1) with comprehensive context and the *time controller* (Figure 5-D2). It allows experts to select a focused period in the stage by moving the slider according to the provided cues, and explore the corresponding details in the *spatiotemporal flow chart*.

The *spatiotemporal flow chart* is based on the Sankey diagram, where the x-axis is a timeline (Figure 5-D3) with the granularity of days and the vertical direction encodes different pollutant patterns in the order of mean AQIs of the pollutant patterns (Figure 5-D4). To reveal the detailed geographical information, we design a *spatial node* (Figure 7), which indicates the cities belonging to the node or the filtered evolution. When hovering on the *spatial node*, it will be enlarged, and the colored points are replaced by two arcs that represent the pollutant patterns of the last day and the next day, respectively. This can help experts trace the evolution of the cities in the node.

The flows in the *spatiotemporal flow chart* represent the evolution among pollutant patterns between two days. The width of flow encodes the number of cities belonging to the evolution. The flows are colored by the linear gradient from the color of the source pollutant pattern to the color of the target pollutant pattern. To facilitate the experts' perception of the relatively important or surprising flows, we propose two metrics, namely *importance* and *surprise*, to set the opacity of flows. *Importance* evaluates the significant degree of evolution. For an evolution flow f_{ij} where i and j are respectively the ranking indexes of the source pollutant pattern and the target pollutant pattern, its importance $S_{Importance}$ is computed as

$$S_{Importance}(f_{ij}) = a \times S_{Span} + b \times S_{Pollution} + c \times N_{Source}, \quad (1)$$

where $S_{Span} = abs(i-j)$, $S_{Pollution} = max(i,j)$, and N_{Source} is the number of cities belonging to the source node. The parameters a, b and c can adjust the focused aspects of *importance*. *Surprise* evaluates the infrequent degree of evolution, and is computed as

$$S_{Surprise} = (N_{Source}/N_{Total} + N_{Target}/N_{Total}) \times S_{Span}, \quad (2)$$

where N_{Target} and N_{Total} are the numbers of cities belonging to the target node and the urban agglomeration respectively. Experts can select one of the two metrics to set the opacity of flows.

We provide various filters to refine flows, which are the filters of flow direction, flow span, city number, flow metric, pollutant pattern and city. The flow-related filters can be set in the *Parameters View* (Figure 5-A2 to A5). The filters of pollutant pattern and city can be set by clicking the cluster glyphs (Figure 5-D5) and the city points in the *spatiotemporal flow chart*, respectively. After setting the filters, the flows that meet all the filters are highlighted.

The *time controller* (Figure 5-D2) comprises three rows. The bottom row contains a time axis representing the stage with the unit of days and a slider that can be dragged on the time axis to select the focused range. The middle row presents the entropy of the occurred pollutant patterns on each day, using colored bars with the *measurement colormap* (Figure 2-C3). It can help experts quickly find the days with consistent or chaotic pollutant patterns. The top row with colored bars presents the flow number of each day when the flow-related filters are set. It helps experts find the days with more filtered flows within the stage. When clicking these colored bars, the focused period will start from the corresponding time step.

Guidance part. To support more effective exploration, we provide guidance based on city communities and frequent patterns. For *city community guidance*, we design a matrix-based view for showing city relationships (Figure 5-E1), whose x-axis and y-axis both encode the cities. The gray line segments (Figure 5-E2) on the axes indicate city communities. The elements of the matrix encode the number of occurrences between the cities and are colored with the *measurement colormap*. A tooltip will be shown when hovering on an element. Moreover, we design the city community card (Figure 5-E3) to show concise information of each city community, which shows the contained cities and the temporal sequence or distribution of pollutant patterns for each city (Figure 5-E3). After selecting a city community card, the related flows are highlighted in the *spatiotemporal flow chart*. Besides, experts can adjust the city communities freely according to their expectations, by clicking the city points in the spatial component of the card. If a highlighted city point is clicked, this city is eliminated from the city community. Otherwise, the clicked city is eliminated from its original community and added into the new city community.

The *frequent pattern guidance* uses the efficient VMSP algorithm [FVWGT14] to extract the common evolution patterns in a stage. To avoid trying different support thresholds repeatedly, we propose a strategy for summarizing the frequent patterns with different support thresholds. Firstly, the extracted frequent patterns are classified by their start pollutant pattern, and the number of patterns in different classes are shown as buttons (Figure 5-E4). Then, the

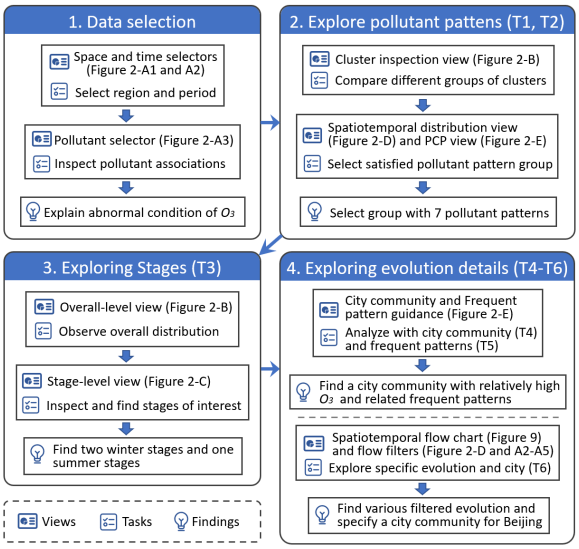

Figure 8: *Workflow of the case study.*

main trends of each class are computed by the CoreFlow algorithm and depicted as a vertical icicle plot (Figure 5-E5) when clicking the corresponding button. Each node in the icicle plot corresponds to a set of frequent patterns that follow the sequence traced from the root node to the current node, which are displayed as the *frequent pattern cards* in Figure 5-E6 after clicking the node. The top of the *frequent pattern card* shows the pattern's support, occurrence count and length. Its left part displays the sequence of pollutant patterns, and the right part is a glyph that shows its spatiotemporal distribution. When selecting a frequent pattern card, the related flows are highlighted in the *spatiotemporal flow chart*.

5 Evaluation

5.1 Case study

Experts E_A and E_B were invited to conduct case studies. To integrate their knowledge from different research fields, experts decided to complete case studies together on one computer where the system was deployed. We recorded their observations and findings by video meeting. Experts operated freely according to their needs without specified tasks. The following describes the case study of Jing-Jin-Ji urban agglomeration with severe air pollution, and the workflow of the case study is shown in Figure 8. Another case study by a new expert E_C is introduced in supplementary materials.

Pollutant association inspection. The experts began with the data selection where all days were selected for Jing-Jin-Ji urban agglomeration. Then, from the *pollutant selector* (Figure 2-A3), E_B noticed that O_3 had the lowest entropy among all the pollutants. This is not in line with his expectation, because the formation of ozone was very unstable and was easily affected by seasons. Experts further examined the links and found that the main pollutants $PM_{2.5}$ and PM_{10} presented high association and both had low associations with O_3. This observation accorded with the inference of E_B. He explained that formations of $PM_{2.5}$ and PM_{10} would consume the precursors of O_3, such as VOCs and nitrogen oxides. Moreover, hazes caused by particulate matter would reduce near-surface light, which also inhibited the formation of O_3. After inspecting the pollutant associations, experts retained all the pollutants.

Exploring pollutant patterns (T1, T2). Experts first had an overall observation of the cluster glyphs in the *cluster inspection view* (Figure 2-B). E_A found that the cluster glyphs with green colors only existed in the rows containing more than 6 clusters and decided to select a cluster group among these finer clustering results. Then, E_A compared the cluster groups with 7 and 8 clusters by inspecting the pollutant distribution, spatiotemporal distribution and PCP view. E_A found that the cluster groups with 7 and 8 clusters are quite similar in both pollutant distribution and spatiotemporal distribution. Thus, they considered that the fineness of the cluster group with 7 clusters could satisfy their need, and selected the cluster group by clicking the icon as shown in Figure 2-B4.

Finding evolution stages of interest (T3). The experts started from the *overall-level view* (Figure 5-B) with stage markers. They found that the overall air quality distribution of Jing-Jin-Ji was consistent with the general knowledge that the air quality in winter was worse because of the urban central heating. E_A also checked the marker positions and thought the stage division was largely reasonable by comparing the colors inside a stage and the colors near the marker. Then, they inspected three stage cards with heavy air pollution in the stage-level view, as shown in Figure 5-C1, C2 and Figure 9-A1. E_A found that the divisions of city communities in these stages were very similar, and one city community had comparatively better air quality and the other had worse air quality. E_B was also interested in another summer stage (Figure 5-C3), where his concerned pollutant pattern with large IAQI of O_3 (Figure 2-B2) most commonly occurred in a city community (Figure 5-C4).

Analyzing city communities (T4) and frequent patterns (T5). E_B further examined the summer stage (Figure 5-C3) and noticed the city relationship view (Figure 5-E1). E_B found that the relationships among the bigger city community were very strong, which implied the concentration of O_3 in these cities might affect each other. Then, E_B selected this city community (Figure 5-E2) and filtered the flows related to the pollutant pattern with a large IAQI of O_3 (Figure 5-D5). From the highlighted flows (Figure 5-D1), he confirmed that the pollutant pattern was dominated among the city community in this summer stage. To further investigate the frequent evolution started from the concerned pollutant pattern, E_B inspected the part of frequent patterns (Figure 5-E5). He found that the concerned pollutant pattern commonly occurred continuously and tended to become the green pollutant pattern. Thus, E_B further observed the frequent patterns started from the green pollutant pattern, and found that evolution commonly turned to the concerned pollutant pattern. After the exploration, the experts had a better understanding of the evolution regarding the concerned pollutant pattern. E_B recorded the stage and the frequent patterns for further research.

Exploring the specific evolution and the focused city (T6). To inspect the winter stages, E_A selected the stage card with the heaviest air pollution (Figure 9-A1). Since the evolution in this stage was complex, E_A first selected the city community with more heavy pollution. Then, E_A noticed the two light colored bars (Figure 9-A2) in the middle row of the *time controller*, which meant the two days had relatively stable air quality among the urban agglomeration. Thus, she checked the two days and found that air quality of all the

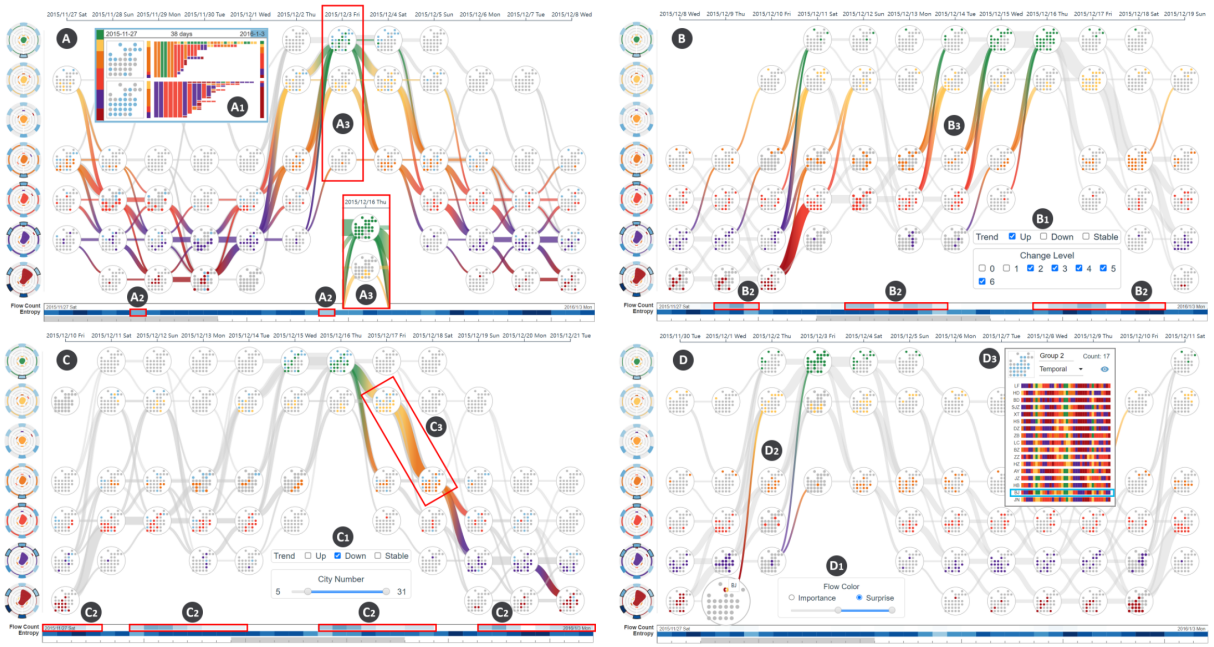

Figure 9: *Exploration of the winter stage. (A) Finding days with low entropy of pollutant patterns. (B) Finding up trends with more than 2 change levels. (C) Finding down trends with more than 5 cities. (D) Finding abnormal evolution with high surprise.*

cities became better in these two days (Figure 9-A3). The experts further filtered the flows with up trend and more than 2 change levels by setting the filters as Figure 9-B1, and found multiple short periods (Figure 9-B2) where the air quality became better. The expert inferred that snowfall might occur in the corresponding cities during these periods. The experts also inspected the evolution with down trends and more than 5 cities (Figure 9-C1), and discovered some periods that existed a large range of air quality deterioration, as shown in Figure 9-C2. The expert clicked some flows of interest (Figure 9-C3) and recorded the corresponding cities. To find the abnormal evolution, the expert switched to the *Surprise* mode and filtered the flows with high surprise values (Figure 9-D1), and found the evolution with drastic changes (Figure 9-D2) in air quality existed on BJ (Beijing). Thus, the experts further selected the city Beijing and observed its evolution. E_B found that the poor air quality in Beijing usually co-occurred with the distant cities. Thus, E_B refined the city community for Beijing as shown in Figure 9-D3, and recorded it for future research.

5.2 Expert interview

After the case studies, we interviewed the domain experts E_A and E_B. Their comments regarding the workflow, visual design and effectiveness of *AirLens* have been recorded. Since the visual designs of our system are also related to the analysis of spatiotemporal data and event sequence, we also invited three researchers $R_A - R_C$ who used the system for the first time. R_A is a post-doctoral scholar focusing on analyzing environmental problems using machine learning and GIS techniques, who is a potential final user of *AirLens*. R_B and R_C are second-year PhD students that have more than two years of experience in the visualization of spatiotemporal data and event sequence data respectively.

Procedure. For the new researchers that had not used *AirLens*, we followed the interview procedure used by Wang et al. [WPL*21]. First, we introduced the data, problem and analysis tasks to them. Second, we demonstrated the workflow, visual designs and interactions through a detailed example. Third, we introduced the process and findings of the case study in Section 5.1. Fourth, the researchers explored the system freely and shared the discovered information of interest with us. Finally, we conducted an interview for each researcher, and collected their feedback and suggestions about the system. The feedback of both domain experts and new researchers is summarized as follows.

System. The domain experts both considered the system effective and comprehensive, which fulfilled all the analysis tasks. Although visual analytics was different from their traditional research methods, they thought the workflow was easy to follow, and the results were credible and explainable because of the comprehensive context in *AirLens*. *"These views can help me understand the metrics and the results of the algorithms"* commented E_B. Moreover, both domain experts appreciated the stage-based analysis and multi-level visualization. *"Our methods generally focused on specific spatial and temporal scales, and time intervals are fixed. The multi-level analysis with stages is flexible, and I can find some interesting conditions that I have not noticed before."* commented E_A.

For the new researchers, they thought that the system was intuitive and flexible. Regarding the *pollutant pattern inspection module*, they considered that it was good to display multiple clustering results at one time. Regarding the *evolution analysis module*, they had a better perception of complex patterns compared with the domain experts. However, they commonly stopped the analysis after focusing on a city and could not propose new hypotheses, due to the

lack of domain knowledge. This also meant that the target users of *AirLens* were mainly environmental experts related to air pollution.

Visual designs and interactions. Both groups of users considered the system was well designed, with informative visualizations and rich interactions. The new researchers could understand most of the visual encodings and interactions after our demonstration. They thought the labels, tooltips and legends provided in the system can help them understand the visualizations. All the users appreciated the stage-level view that presented the general evolution trends of different city communities. *"The general trends for different city communities are shown clearly"* (E_A). *"I can easily compare two stages, as well as the city communities in a stage"* (R_B). Most users thought the flow filters were very useful to find the desired information. Especially, the filters were linked with the *time controller*, which was appreciated by R_A and R_C. R_A and R_B also appreciated the visual designs of different kinds of cards, which can present rich information and remind users to click them. Besides, R_C thought the exploration of frequent patterns computed by different support thresholds was very useful, since she had been confused with the setting of the support threshold of VMSP algorithm.

Some problems were also commented by the researchers. R_A and R_C considered some pollutant pattern colors of the same pollution level were difficult to distinguish. E_A and R_B commented that the city relationship view was inconvenient to find the relationship between two specific cities, since city names were not shown directly.

Suggestions. Both groups of users also provided valuable suggestions. From the perspective of system functions, domain expert E_A advised us to support the analysis history, thus they could continue with the previous analysis quickly. E_B suggested the function of saving the findings of interest as a structured file, which can reduce his burden of recording the findings. From the perspective of visual design, R_A suggested a separate comparison view for comparing different urban agglomerations. R_B commented that more statistical information could be provided for the focused city.

6 Discussion

6.1 System performance

Since *AirLens* aims to support online analysis according to the selected space, time and pollutants, we adopt efficient algorithms for the analysis tasks, including K-Means, Louvain and VMSP. Their average computation time for the largest urban agglomeration (41 cities) with the longest period (364 days) are as follows: clustering 8.97 seconds, stage division 0.74 seconds, city community division 0.59 seconds, Coreflow 0.21 seconds, and VMSP 2.82 seconds. It can be shown that the most time-consuming process is clustering, which means experts only need to wait at first and can explore the data fluently during the subsequent analysis. Our system has the potential to be applied to larger datasets, since the computation processes of clustering and frequent pattern mining are easy to be parallelized, which can largely reduce the computation time for the data with very large spatial and temporal scales.

6.2 Generalizability

First, the pipeline, methods and most of the designs can be applied to the grid-based air quality data generated by simulation models. The space-related designs can be easily changed to display grids. Second, the *pollutant pattern inspection module* can be used to summarize and visualize other multivariate spatiotemporal data, such as climate data and vegetation change data, where domain experts also need to explore typical multivariate patterns. Finally, the exploration process of frequent patterns can be applied to general event sequence data, such as learning behavior data and medical record data, for exploring patterns with different frequent degrees.

6.3 Limitations

Although the effectiveness of *AirLens* has been evaluated by the case study and expert interview, it has several limitations that need to be improved. The first concern is the scalability of our visual designs. Although we have designed multi-level visualizations for the long-period air quality data, the scalability problem still exists when the number of pollutant patterns is large. On the one hand, the height of the *detail-level view* is proportional to the number of pollutant patterns. Although we have used an adaptive strategy to set the height of cluster glyphs, the limited screen space cannot support many pollutant patterns. On the other hand, the color encoding of pollutant patterns faces a scalability problem. Although we have used a sequential colormap for each standard color in the *IAQI color scheme*, it is also hard to distinguish multiple patterns that are in the same pollution level, when the number of pollutant patterns is large. In *AirLens*, the numbers of pollutant patterns are set from 4 to 10, which is considered as sufficient by domain experts. For supporting more pollutant patterns, we plan to add the pattern index labels to the corresponding visual elements along with the colors.

The other limitation is that *AirLens* mainly helps experts discover significant information or verify their hypothesis, but cannot support further reasoning the discovered information, due to lacking the data of air quality influence factors. More datasets can be integrated into *AirLens* for supporting further reasoning.

7 Conclusions

In this paper, we propose *AirLens*, a comprehensive visual analytics system that supports domain experts in exploring the complex air quality evolution in urban agglomeration through two analysis modules. The *pollutant pattern inspection module* helps experts summarize the typical pollutant patterns by combining the automatic algorithm with experts' knowledge, which facilitates experts' understanding and reduces the perception complexity. On this basis, the *evolution analysis module* helps experts explore the air quality evolution from multiple levels, including the overall level, stage level and detail level. Moreover, with the help of frequent pattern mining, city community extraction and various filters, experts can understand the air quality evolution among the urban agglomeration and discover significant information from multiple aspects. The comprehensive case study and the expert interview have evaluated the effectiveness and usability of our system.

In the future, we intend to improve the system according to the feedback, including promoting the scalability of visual designs and providing more information for the focused city. To further help experts conduct in-depth quantitative analysis and reason the discovered information, we plan to incorporate more datasets and develop a new module that helps experts explore the relationships between air quality and its influence factors.

Acknowledgements

This research was supported by National Natural Science Foundation of China under Grant 42171450 and Key R&D Project of Science and Technology Development Plan of Jilin Province under Grant 20210201074GX.

References

[AAFW17] ANDRIENKO G., ANDRIENKO N., FUCHS G., WOOD J.: Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data. *IEEE Transactions on Visualization and Computer Graphics 23*, 9 (2017), 2120–2136. 2

[AFGY02] AYRES J., FLANNICK J., GEHRKE J., YIU T.: Sequential pattern mining using a bitmap representation. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2002), KDD '02, Association for Computing Machinery, pp. 429–435. 2

[BDM*17] BEECHAM R., DYKES J., MEULEMANS W., SLINGSBY A., TURKAY C., WOOD J.: Map lineups: Effects of spatial structure on graphical inference. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 391–400. 2

[BG05] BORG I., GROENEN P. J.: *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005. 4

[BGLL08] BLONDEL V. D., GUILLAUME J.-L., LAMBIOTTE R., LEFEBVRE E.: Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment 2008*, 10 (2008), P10008. 6

[BS06] BYUN D., SCHERE K. L.: Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. *Applied Mechanics Reviews 59*, 2 (03 2006), 51–77. 2

[CT90] COVER T., THOMAS J.: *Elements of information theory. 2nd ed.* 11 1990. 4

[CXR18] CHEN Y., XU P., REN L.: Sequence synopsis: Optimize visual summary of temporal event data. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (2018), 45–55. 2

[CYP*20] CHEN Q., YUE X., PLANTAZ X., CHEN Y., SHI C., PONG T.-C., QU H.: Viseq: Visual analytics of learning sequence in massive open online courses. *IEEE Transactions on Visualization and Computer Graphics 26*, 3 (2020), 1622–1636. 2

[DWC*19] DENG Z., WENG D., CHEN J., LIU R., WANG Z., BAO J., ZHENG Y., WU Y.: Airvis: Visual analytics of air pollution propagation. *IEEE transactions on visualization and computer graphics 26*, 1 (2019), 800–810. 1, 2

[EKSX96] ESTER M., KRIEGEL H.-P., SANDER J., XU X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (1996), KDD'96, AAAI Press, pp. 226–231. 4

[EPD16] EPD: *Technical Regulation on Ambient Air Quality Index (AQI) (on Trial)*. Tech. rep., 2016. 3, 4

[Fox91] FOX W. R.: Finding groups in data: An introduction to cluster analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 40*, 3 (1991), 486–487. 4

[FVWGT14] FOURNIER-VIGER P., WU C.-W., GOMARIZ A., TSENG V. S.: Vmsp: Efficient vertical mining of maximal sequential patterns. In *Advances in Artificial Intelligence* (Cham, 2014), Sokolova M., van Beek P., (Eds.), Springer International Publishing, pp. 83–94. 2, 7

[FWX*16] FAN X., WANG L., XU H., LI S., TIAN L.: Characterizing air quality data from complex network perspective. *Environmental science and pollution research international 23*, 4 (February 2016), 3621–3631. 2

[GGJ*21] GUO Y., GUO S., JIN Z., KAUL S., GOTZ D., CAO N.: A survey on visual analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. 2

[GJG*19] GUO S., JIN Z., GOTZ D., DU F., ZHA H., CAO N.: Visual progression analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 417–426. 2

[GPS*05] GRELL G. A., PECKHAM S. E., SCHMITZ R., MCKEEN S. A., FROST G., SKAMAROCK W. C., EDER B.: Fully coupled "online" chemistry within the wrf model. *Atmospheric Environment 39*, 37 (2005), 6957–6975. 2

[GRPF16] GRUENDL H., RIEHMANN P., PAUSCH Y., FROEHLICH B.: Time-series plots integrated in parallel-coordinates displays. *Computer Graphics Forum 35*, 3 (2016), 321–330. 2

[GTC*19] GUO F., TIANLONG G., CHEN W., WU F., WANG Q., SHI L., QU H.: Visual exploration of air quality data with a time-correlation-partitioning tree based on information theory. *ACM Transactions on Interactive Intelligent Systems 9* (02 2019), 1–23. 1

[GXZ*18] GUO S., XU K., ZHAO R., GOTZ D., ZHA H., CAO N.: Eventthread: Visual summarization and stage analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (2018), 56–65. 2, 6

[GYL*17] GUODAO S., YAJUAN H., LI J., XIAORUI J., RONGHUA L.: Urban agglomerations-based visual analysis of air quality data. *Journal of Computer-Aided Design & Computer Graphics 29*, 1 (2017), 3. 2

[HLZ*21] HAN J., LIU H., ZHU H., XIONG H., DOU D.: Joint air quality and weather prediction based on multi-adversarial spatiotemporal networks. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021* (2021), AAAI Press, pp. 4081–4089. 2

[Kra08] KRAAK M.-J.: The space-time cube revisited from a geovisualization perspective. *Proc 21st Int Cartogr Conf* (07 2008). 2

[LC06] LEE E. M., CHAN K. C.: Discovering association patterns in large spatio-temporal databases. In *Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)* (2006), pp. 349–354. 1

[LCCC17] LI X., CHENG Y., CONG G., CHEN L.: Discovering pollution sources and propagation patterns in urban area. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017* (2017), ACM, pp. 1863–1872. 1

[LCZ*19] LI J., CHEN S., ZHANG K., ANDRIENKO G., ANDRIENKO N.: Cope: Interactive exploration of co-occurrence patterns in spatial time series. *IEEE Transactions on Visualization and Computer Graphics 25*, 8 (2019), 2554–2567. 2

[LKD*17] LIU Z., KERR B., DONTCHEVA M., GROVER J., HOFFMAN M., WILSON A.: Coreflow: Extracting and visualizing branching patterns from event sequences. *Computer Graphics Forum 36*, 3 (2017), 527–538. 2, 6

[LVG*21] LIU D., VEERAMACHANENI K., GEIGER A., LI V. O. K., QU H.: Aqeyes: Visual analytics for anomaly detection and examination of air quality data. *CoRR abs/2103.12910* (2021). arXiv:2103.12910. 2

[LWSY20] LIU C., WU C., SHAO H., YUAN X.: Smartcube: An adaptive data management architecture for the real-time visualization of spatiotemporal datasets. *IEEE Transactions on Visualization and Computer Graphics 26*, 1 (2020), 790–799. 2

[LXZ*16] LI J., XIAO Z., ZHAO H., MENG Z., ZHANG K.: Visual analytics of smogs in china. *Journal of Visualization 19*, 3 (2016), 461–474. 1

[LZM14] LI J., ZHANG K., MENG Z.-P.: Vismate: Interactive visual analysis of station-based observation data on climate changes. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2014), pp. 133–142. 2

[Mac67] MacQueen J.: Some methods for classification and analysis of multivariate observations. vol. 1, pp. 281–297. 4

[MGDVT18] Marinoni A., Gamba P., De Vecchi D., Tuia D.: Discovering temporal patterns of air quality in different parts of europe with data driven feature extraction. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* (2018), pp. 2062–2065. 1

[PW14] Perer A., Wang F.: Frequence: Interactive mining and visualization of temporal frequent event sequences. In *Proceedings of the 19th International Conference on Intelligent User Interfaces* (New York, NY, USA, 2014), IUI '14, Association for Computing Machinery, pp. 153–162. 2

[QCX*07] Qu H., Chan W.-Y., Xu A., Chung K.-L., Lau K.-H., Guo P.: Visual analysis of the air pollution problem in hong kong. *IEEE Transactions on visualization and Computer Graphics 13*, 6 (2007), 1408–1415. 1

[QDL*21] Qiao H., Deng Z., Li H., Hu J., Song Q., Xia C.: Complex networks from time series data allow an efficient historical stage division of urban air quality information. *Applied Mathematics and Computation 410* (2021), 126435. 2

[QLKL19] Qi Y., Li Q., Karimian H., Liu D.: A hybrid model for spatiotemporal forecasting of pm2.5 based on graph convolutional neural network and long short-term memory. *Science of The Total Environment 664* (2019), 1–10. 2

[QLR*20] Qu D., Lin X., Ren K., Liu Q., Zhang H.: Airexplorer: Visual exploration of air quality data based on time-series querying. *Journal of Visualization 23*, 6 (2020), 1129–1145. 2

[RWZ*20] Ren K., Wu Y., Zhang H., Fu J., Qu D., Lin X.: Visual analytics of air pollution propagation through dynamic network analysis. *IEEE Access 8* (2020), 205289–205306. 2

[SDR*15] Stein A. F., Draxler R. R., Rolph G. D., Stunder B. J. B., Cohen M. D., Ngan F.: Noaa's hysplit atmospheric transport and dispersion modeling system. *Bulletin of the American Meteorological Society 96*, 12 (2015), 2059 – 2077. 2

[Sha48] Shannon C.: A mathematical theory of communication. *Bell System Technical Journal 27* (01 1948), 379–423. 4

[SMM12] Sedlmair M., Meyer M., Munzner T.: Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis) 18*, 12 (2012), 2431–2440. 3

[SZF*19] Shen Y., Zhang L., Fang X., Ji H., Li X., Zhao Z.: Spatiotemporal patterns of recent pm2.5 concentrations over typical urban agglomerations in china. *Science of The Total Environment 655* (2019), 13–26. 1, 2

[TC05] Thomas J. J., Cook K. A.: *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005. 2

[WAM01] Weber M., Alexa M., Muller W.: Visualizing time-series on spirals. In *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001.* (2001), pp. 7–13. 2

[WGW*20] Wu J., Guo Z., Wang Z., Xu Q., Wu Y.: Visual analytics of multivariate event sequence data in racquet sports. In *2020 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2020), pp. 36–47. 2, 3

[WLG*21] Wu J., Liu D., Guo Z., Xu Q., Wu Y.: Tacticflow: Visual analytics of ever-changing tactics in racket sports. *IEEE Transactions on Visualization and Computer Graphics* (2021). 2

[WLS*21] Wang Y., Liang H., Shu X., Wang J., Xu K., Deng Z., Campbell C. D., Chen B., Wu Y., Qu H.: Interactive visual exploration of longitudinal historical career mobility data. *IEEE Transactions on Visualization and Computer Graphics* (2021). 2

[WPL*21] Wang Y., Peng T.-Q., Lu H., Wang H., Xie X., Qu H., Wu Y.: Seek for success: A visualization approach for understanding the dynamics of academic careers. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. 9

[WSH13] Wang J., Sisneros R., Huang J.: Interactive selection of multivariate features in large spatiotemporal data. In *2013 IEEE Pacific Visualization Symposium (PacificVis)* (2013), IEEE, pp. 145–152. 5

[WWS*16] Wu T., Wu Y., Shi C., Qu H., Cui W.: Piecestack: Toward better understanding of stacked graphs. *IEEE Transactions on Visualization and Computer Graphics 22*, 6 (2016), 1640–1651. 2

[ZCWY14] Zheng Y., Capra L., Wolfson O., Yang H.: Urban computing: Concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol. 5*, 3 (sep 2014), 38:1–55. 2

[ZDZZ21] Zhu J., Deng F., Zhao J., Zheng H.: Attention-based parallel networks (apnet) for pm2.5 spatiotemporal prediction. *Science of The Total Environment 769* (2021), 145082. 2

[ZLD*15] Zhao J., Liu Z., Dontcheva M., Hertzmann A., Wilson A.: Matrixwave: Visual comparison of event sequence data. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY, USA, 2015), CHI '15, Association for Computing Machinery, pp. 259–268. 2

[ZRL*19] Zhang H., Ren K., Lin Y., Qu D., Li Z.: Airinsight: Visual exploration and interpretation of latent patterns and anomalies in air quality data. *Sustainability 11*, 10 (2019), 2944. 2

[ZYL*15] Zheng Y., Yi X., Li M., Li R., Shan Z., Chang E., Li T.: Forecasting fine-grained air quality based on big data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2015), KDD '15, Association for Computing Machinery, pp. 2267–2276. 2

[ZYL*17] Zhou Z., Ye Z., Liu Y., Liu F., Tao Y., Su W.: Visual analytics for spatial clusters of air-quality data. *IEEE Computer Graphics and Applications 37*, 5 (2017), 98–105. 1, 2