

Jie Li · Zhao Xiao · Han-Qing Zhao ·  
Zhao-Peng Meng · Kang Zhang

## Visual analytics of smogs in China

Received: 12 August 2015 / Accepted: 8 December 2015 / Published online: 5 January 2016  
© The Visualization Society of Japan 2015

**Abstract** Smog is one of the most important environmental problems in China. Scientists have attempted to explain the causes from chemistry, physics, atmosphere and other perspectives. Many meteorologists believe that meteorology is a crucial reason. In this paper we present a new multi-view approach to visual analytics of the recent smog problems in China. This approach integrates four interrelated visualizations, each specialized in a different analysis task. To reveal the relationship between smog and meteorological attributes, we design a Correlation Detection View that simultaneously visualizes the air quality change patterns across multiple cities and related meteorological attributes. Component Trend View is used to show the variation patterns of six air components, while Aggregation View can reveal the regional overall pollution situations. A case study has been conducted using the China Air Quality Observation Data and European Centre for Medium-Range Weather Forecasts re-analysis data to verify the effectiveness of the proposed approach. By visually analyzing the meteorological data of Beijing and other major cities, we have found several interesting patterns, which prove the validity of our work in identifying sources of smog.

**Keywords** Smog · Spatiotemporal visualization · Meteorological visualization · Visual analytics

### 1 Introduction

Recently, smog has become one of the most serious problems in China, which has attracted nationwide attention (Ma et al. 2012). Many cities are immersed in smog and the blue skyline has been barely visible over the past few years. Smog not only affects air quality, but also brings a series of negative effects on almost every aspect of our life. It can cause all kinds of respirational diseases (Chen et al. 2013), increased transportation risks, and reduced urban sustainable competitive advantages (Chen et al. 2014). To tackle the

---

J. Li · Z. Xiao · H.-Q. Zhao · Z.-P. Meng  
School of Computer Software, Tianjin University, Tianjin, China  
E-mail: jie.li@tju.edu.cn

Z. Xiao  
E-mail: zxiao@tju.edu.cn

H.-Q. Zhao  
E-mail: hanqingzh@tju.edu.cn

Z.-P. Meng  
E-mail: zpmeng@tju.edu.cn

K. Zhang (✉)  
Department of Computer Science, University of Texas at Dallas, MS EC31, Box 830688,  
Richardson, TX 75083-0688, USA  
E-mail: kzhang@utdallas.edu

smog problem, researchers have conducted various studies on air pollution, and both the state and local governments have legislated the related laws and regulations. The pollution situations, however, are still serious, and the real pollution sources for major cities, such as Beijing, have not been accurately detected.

Due to our limited understanding on the causes of smog, all the analytical results of the existing methods are with inherent uncertainties (Zhou et al. 2013), and even conflicting conclusions were often drawn by different researchers and organizations. For example, once an official claimed that smog was mainly caused by the smoke of cooking, but Beijing government considered the vehicle exhaust as the main cause of smog. They tried to reduce car exhaust gases by a “Private Vehicle Restriction” policy and limiting the vehicles from other provinces to enter the city. However, smog in Baoding (a small city near Beijing) is heavier than in Beijing, yet Baoding has far fewer vehicles than Beijing, making one to doubt about such measures. In fact, although the discharge amount of pollution gases in a city is in a stable state, the air quality of the city has always been changing affected by meteorological factors, such as atmosphere movement and inversion layer (the air temperature at a low altitude is lower than that at a high altitude at the same location). It is impossible to analyze smog in a city without considering the meteorological attributes of its surrounding area (Hulek et al. 2013).

We thus attempt to design a visualization framework that can intuitively and interactively reveal the evolutionary nature of air quality at different spatiotemporal scales. Our goal is to identify spatiotemporal patterns of air quality changes and the relations between air quality and meteorological attributes through visual analysis. The major contributions of this paper include:

- A comprehensive visual analytics framework for discovering air pollution patterns at different spatiotemporal scales.
- Three visual analytic views for different analysis tasks in smog studies.
- Case studies on China Air Quality Observation Data and ECMWF re-analysis data, through which the effectiveness and the scalability of the approach have been verified.

The remaining part of this paper is organized as follows: Sect. 2 reviews the related work. The approach overview is then described in Sect. 3, followed by three concrete visualization views in Sect. 4. Section 5 describes a case study, and an expert review is described in Sect. 6. Finally, we conclude the paper in Sect. 7.

## 2 Related work

Our current work builds upon our previous system that has been developed to explore and visualize oceanographic applications in a distributed environment (Li et al. 2014a, b). In this section, we first discuss the existing spatiotemporal data visualization techniques, and then review several related environment visualization applications.

### 2.1 Spatiotemporal data visualization

Thematic map (Slocum et al. 2009), which can be viewed as an overlay of Heatmap or glyphs on a map, is perhaps the most traditional method of geo-related data visualization. Taggram (Nguyen and Schumann 2010) is another type of thematic map, in which tag clouds (Lee et al. 2010) are plotted on a map to represent the area characteristics. These methods work well for static data, but provide limited supports for visualizing time-oriented geographic data. Many researchers have attempted to overcome this problem by combining a map with other time-series visualization techniques. Malik et al. (2012) comprehensively used maps, bar charts, line charts and pie charts to analyze the correlations between urban crime activities and spatiotemporal dimensions. Landesberge et al. (2012) designed a dynamic categorical data view (DCDV) to visualize human position transitions in 1 day, associated it with a geography view. Furthermore, parallel coordinates (Johansson and Jern 2007; Tominski et al. 2004), ThemeRiver (Havre et al. 2000, 2002), stacked bar charts (Nocke et al. 2004), and many other existing time-series visualization techniques can be combined with a map to form effective spatiotemporal data visualization tools.

**Table 1** Attributes of the collected air quality data

ID	Name	Unit
1	AQI (Air Quality Index)	Computed based on 2–7
2	CO	mg/m <sup>3</sup>
3	NO <sub>2</sub>	ug/m <sup>3</sup>
4	SO <sub>2</sub>	ug/m <sup>3</sup>
5	O <sub>3</sub>	ug/m <sup>3</sup>
6	PM2.5	ug/m <sup>3</sup>
7	PM10	ug/m <sup>3</sup>

## 2.2 Visual analysis environment data

Visualization has always been an effective tool in environmental studies, and many classic visualization methods exist, such as contour line (Watson 1992; Johansson et al. 2010), standard coloration (Li et al. 2014a, b), etc. These methods can only show analysis results, lacking interactions and the abilities for discovering the potential knowledge from the huge amounts of meteorological data. Johansson et al. (2010) used a 3D GIS platform to show the meteorological data. Both Yannier et al. (2008) and Janicke et al. (2009) have performed the studies of weather variation visualization, similar to our previous work. Their studies, however, focus on the use of touch screens in enhancing the perception effects.

The most related idea was found in Qu et al. (2007), which analyses the air pollution problem in Hong Kong. Several visualization tools have been proposed, such as S-shape parallel coordinates, polar system with circular pixel bar and weighted complete graph. However, small scale analysis cannot prove the effectiveness of the method. Our method focuses on the smog problem at a large scale. We also analyze different city clusters in multiple latitudes.

## 3 Approach overview

### 3.1 Data source

China's Ministry of Environmental Protection and Environmental Monitoring Centre Station has operated a national air quality observation network containing about thousands of observation stations throughout the country. These stations can continuously observe and record air qualities at a fixed temporal interval and automatically transfer these data to the specified data centers. Through an online data service, we could obtain air quality observation data of 846 stations in each hour. The attributes of air quality data are summarized in Table 1.

To analyse the correlations between the smog of a city and the meteorological attributes in the surrounding areas of the city, we also need to collect meteorological data containing multiple parameters. The ECMWF re-analysis data with the spatial resolution of  $1^\circ \times 1^\circ$  is used in our approach.

An inversion means that the air temperature at the lower altitude is lower than that at the higher altitude. An inversion layer forms a stable air structure and makes it difficult for smog to dissipate. Wind field and inversion layer are two major factors to the formation of smog, therefore wind and air temperature are the two parameters in the downloaded grid data.

### 3.2 Objectives

Among all the tasks, understanding the overall dynamics in different spatiotemporal scales is the most important. For example, “which is the heaviest smog area?”, and “whether the smog problem is a regional or a nationwide phenomenon?” Apart from the overall dynamics, environmental experts also wish to know the exact pollution sources in major cities to guide the government in developing economic and industrial programs. Based on the “Beijing–Tianjin–Hebei integration strategy”, all the high-polluting enterprises in Beijing should be transferred to other places and the new locations should be fully evaluated to minimize the impacts on Beijing resulting from air movement. Understanding the temporal trend of smog at different spatiotemporal scales becomes an important task for the above strategy. Experts wish to know not only the overall smog changes in a year or a month, but also a city's air components in each hour to accurately detect the pollutant's origins. Ideally, they wish to be able to forecast the air quality, which is closely related to the

public life. To better design the visualization framework, we classify the most important tasks of smog studies into three major categories:

- Overall distribution: identifying the air pollution situation at different spatial scales.
- Correlation analysis: discovering the impacts of meteorological attributes on air quality, and even predicting the future change based on the patterns observed today.
- Temporal trend: determining the air pollution change in any temporal interval, and the variations of air pollution components at different temporal scales.

### 3.3 Visualization views

Based on the previous system architecture (Li et al. 2014a, b), we add a new visual analysis layer, in which four views, specialized in the different analysis objectives, are integrated.

- Global Distribution View—a radial layout plot including a map that shows the global spatial distribution, a sector-based band for displaying temporal trend, and several clustering rings, in which the K-Means algorithm is adapted to cluster all the cities into groups of similar smog change trends.
- Correlation Detection View—visualizing air quality and meteorological data in one framework. Users can simultaneously explore two types of data via interactions, and analyze the impacts of different meteorological parameters on cities' air quality changes.
- Component Trend View—depicting the daily and hourly air component changes of all the cities. The cities are projected onto a radial scatterplot (Astrolabium) containing multiple axes, each representing one type of pollutant, such as SO<sub>2</sub>, NO<sub>2</sub>, etc. When a city is selected by clicking on a point in the Astrolabium, the detailed hourly component changes in different temporal intervals for the city are shown on an Air Component River.
- Aggregation View—supporting exploration and comparison of the air quality in different cities or regions within a 2D/3D combined context. The aggregation functions often used in DBMS are included in this view, which allows the user to select interested attributes to be visually explored.

All the four views are interrelated. As the main visualization, the Global Distribution View displays both the spatial, temporal data, from which other views' input parameters can be selected by through mouse interaction. In addition, the cities and temporal intervals analysed in other views are automatically highlighted in the Global Distribution View.

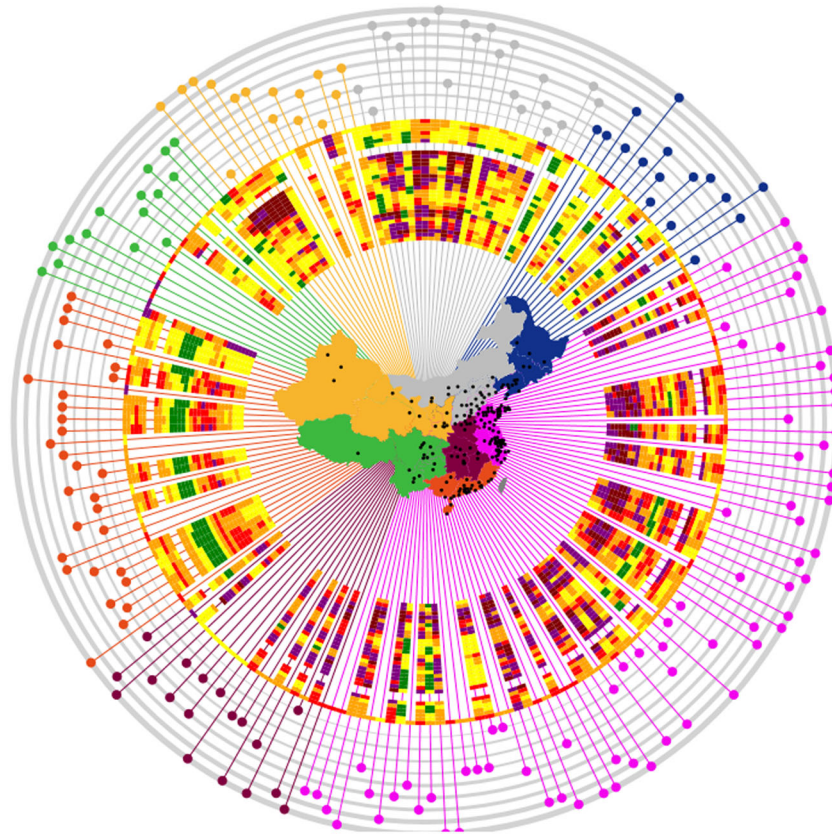
## 4 Multi-view visualization

We first introduce the Correlation Detection View specialized in analysing the fluctuations of smog in a city with the variations of different meteorological attributes.

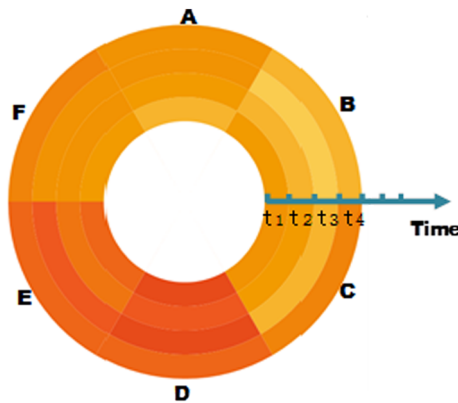
### 4.1 Global Distribution View

To visually encode air quality data, we first use the radial map (Li et al. 2014a, b) containing three visual elements: (1) a map in the center conveying the spatial information; (2) a ring band outside the map, encoding temporal trend changes in the user-defined interval, and (3) multiple concentric rings outside the ring band representing the different clusters generated by a clustering algorithm using the air quality changes rates as the clustering reference, as illustrated in Fig. 1. Using such three components, the spatial, temporal and the clustering information of all the cities can be jointly visualized in a concise and intuitive manner.

All the cities are drawn in the innermost map to illustrate the spatial distribution. Furthermore, each city takes up a sector. To keep the symmetry of the view, the angle interval of each city's sector is equal. For  $N$  cities, the angle of each sector is  $360/N$ . A colored line is drawn along the radial direction, through a sector-based band, and finally locates on a concentric ring. Each small bins in the sector-based band represents the attribute value of the corresponding city at a time point, and the entire band can reflect the change trend in the defined temporal interval, as illustrated in Fig. 2. Outside the sector-based band, each concentric ring indicates a cluster of cities that share similar temporal change rates. We introduce the concrete design of the view in three aspects:



**Fig. 1** Global Distribution View consists of a map, a ring band outside the map, and multiple concentric rings, showing the clustering information



**Fig. 2** Temporal mapping represented by a sector-based ring-band

#### 4.1.1 Spatial mapping

By aligning the cities (with 3D coordinates, i.e. longitude, latitude, altitude) to a 1D angular coordinate system, we can create an extra space for visualizing time and clustering information. To minimize the loss of spatial information when computing the angular coordinate for each city, we first divide China into 7 subareas according to the traditional geographic convention (see the different colors in Fig. 1), and assign an angular span to each subarea in proportion to the number of the cities in that area. The cities in one subarea

are placed in descending order of their latitudes. Our method does not limit the sorting reference. Other parameters, such as longitude, altitude, etc., can also be selected upon application needs.

The cities are arranged radially on different concentric rings along the circumference of a circle at equal angular intervals. This arrangement aims to effectively condense a large number of widely distributed cities in a single view, while emphasizing the map as the core of the visualization.

To keep the map style intuitive and minimize the viewer's cognitive effort, we use the online tool ColorBrewer (Brewer and Harrower 2009) to assign a color in the recommended color solution to each geographical subareas, while the cities are colored the same as the geographical subarea they belong to.

#### 4.1.2 Temporal mapping

The radial direction represents the time axis. Each sector indicates the time series of a city, while a radial bin is colored to represent the concrete AQI (Air Quality Index) value at an instance of time. Figure 2 shows an example of time series visualization of at six cities in 4 years.

#### 4.1.3 Clustering

To represent the clustering information, several concentric rings are used. Each ring represents one cluster, and a ring's thickness represents the absolute value of the average slope of linear regression line between time and the AQI. The cities in a cluster are drawn on the corresponding ring and are highlighted when the mouse moves on the ring.

To support clustering, we first compute the slope to represent the climate change condition of a station. Let  $X = \{x_1, x_2, \dots, x_N\}$  be the set of time instances in the selected time interval, and  $Y = \{y_1, y_2, \dots, y_N\}$  be the set of AQI values at the corresponding instances of time. The slope is calculated as follows:

$$\text{slope} = \frac{\sum_{i=1}^N ((x_i - \bar{x})(y_i - \bar{y}))}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (1)$$

We adapt the K-Means algorithm to generate clusters of slopes and map each cluster onto a ring. Of course, our framework does not depend on any specific clustering algorithm. An extra merit of using the K-Means algorithm is that average slope of each cluster can also be obtained when performing the algorithm.

## 4.2 Correlation Detection View

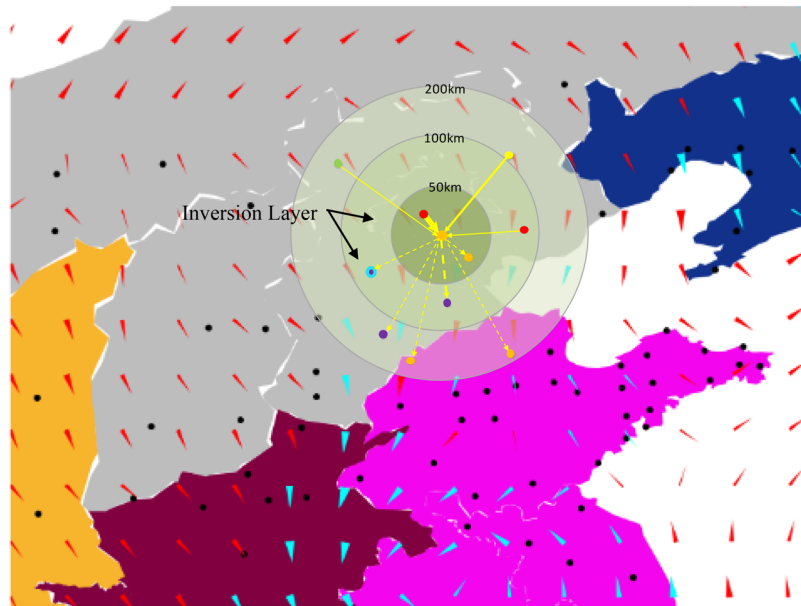
The air quality of a city has always been changing and affected by its surrounding areas due to air movement. We introduce a Correlation Detection View specialized in analyzing the fluctuations of smog in a city with the variations of different meteorological attributes. The Correlation Detection View is constructed based on our previous oceanographic data visualization system (Li et al. 2014a, b). We add a smog correlation detection component which is similar to a radar plot, as shown in Fig. 3.

#### 4.2.1 Meteorological factor visualization

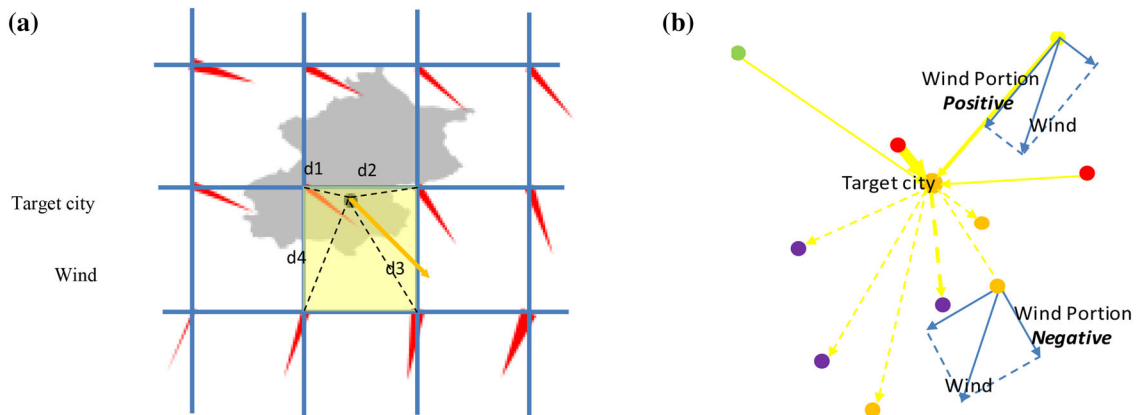
We use the ECMWF re-analysis data as the input. Because the format of such data has been supported in our previous system, we can conveniently integrate the downloaded meteorological data into our framework. The Correlation Detection View visualizes the wind field and the inversion layer, which are used as the context of analyzing the correlation between meteorological factors and smog.

A wind record at a grid point consists of a horizontal component and a vertical component. By calculating the vector sum of such two components, we can get the speed and the direction of wind at that grid point. We use a long equilateral triangle to point to the wind direction, and the width of the triangle's bottom encodes the wind speed. To determine whether there exists an inversion layer at a grid point, we compare the temperatures of different altitudes at the grid point. If an inversion layer exists, the triangle at that point is highlighted in a different color, as shown in Fig. 3.

Through the case study (see Sect. 5), we discover that inversion layers and wind fields are highly related to serious smog phenomena.



**Fig. 3** Correlation Detection View

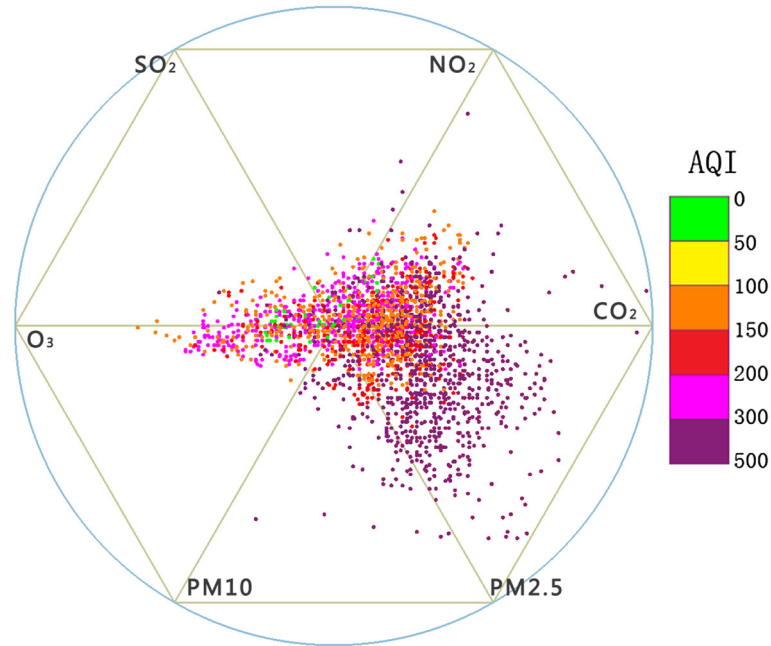


**Fig. 4** **a** Computing a city’s meteorological data based on its four closest grid points. The *yellow arrowed line* is the wind at the target city. **b** Decomposing the wind at neighboring cities to obtain wind portions directed at the target city

#### 4.2.2 Correlation analysis

Correlation analysis aims at identifying the air quality of a city affected by its neighboring cities. Because the cities with air quality observation records may not be exactly on the grid points, we first calculate the wind field (speed and direction) and air temperature of each city using an interpolation algorithm. The four nearest grid points around the city are used as the inputs to the interpolation algorithm, while the weighting of each point is inversely proportional to the distance between the city and the point, as shown in Fig. 4a.

Having obtained the meteorological data and air quality observation data for every city, we select a city as the *target city* (see Fig. 4b) to analyze the relationship of the smog movement between the city and its surrounding cities. For each neighboring city located within the spatial range of this city, we compute the wind portion (see Fig. 4b) that originates from a neighboring city and points to the target city (or away from the target city), by decomposing the wind using the parallelogram law.



**Fig. 5** Component Astrolabium

Let  $D = \{d_1, d_2, \dots, d_N\}$  be the set of distances between the target city and its  $N$  neighboring cities, and  $V = \{v_1, v_2, \dots, v_N\}$  be the set of the wind portions of the neighboring cities. The correlation  $c_i$  between the  $i$ th city and the target city can be simply represented as:

$$c_i = \frac{d_i}{|v_i|} \quad (2)$$

The correlation is drawn as an arrowed line. The solid line indicates that the air quality of the target city is affected by the other cities, while the dotted line reflects that the target city affects its neighboring cities (the wind blew from the target city to these cities). The thicker, the arrowed line is, the stronger the correlation between two cities.

The color of each city (see the small circles in Fig. 3) represents the air quality, with the legend shown in Fig. 5. If there exists an inversion layer above a city, the city's outline is highlighted (Fig. 3). The neighboring cities that have both a strong correlation with the target city and poor air quality should attract our attention during analysis. Our approach also provides the filtering control to help the user to view only the cities that satisfy the defined correlation and air quality conditions.

### 4.3 Component Trend View

Identifying the primary pollutants and variation patterns of different air pollutants is the prerequisite for preventing and controlling smog. We therefore design a Component Trend View consisting of a Components Astrolabium and an Air Component River, based on an Astrolabe method (Wang et al. 2013), RadViz (Sharko et al. 2008) and ThemeRiver (Havre et al. 2002).

#### 4.3.1 Component Astrolabium

A Component Astrolabium contains several axes, each representing a type of pollutant contained in the air quality observation data. We first normalize all the pollutant values into the range between 0 and 1. A Cartesian coordinate system is then constructed, using the center of the Astrolabium as the origin. The air quality of a city at a time can be viewed as an set  $Q = \{pollutant_1, pollutant_2, \dots, pollutant_N\}$ , ( $N$  is the number of pollutant types), and each component in the set  $Q$  has a 2D coordinate. Let  $(x, y)$  be the coordinate of a city, and  $(u_i, v_i)$  be the component on the  $i$ th axis. We can calculate the coordinate of each



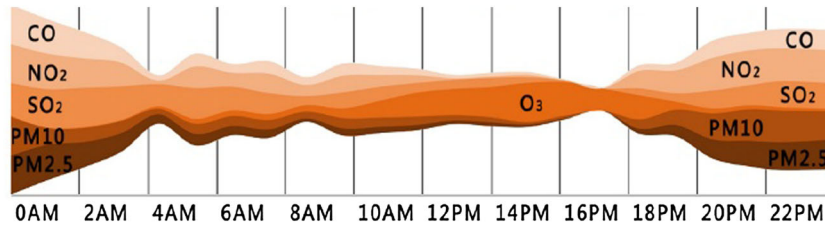


Fig. 6 Air Component River

city by simply computing the vector sum of 6 pollutants (the attribute value on each axis) of the city using Eq. 3:

$$\begin{cases} pollutant_i = (u_i, v_i) \\ (x, y) = \sum_{i=0}^N (u_i, v_i) \end{cases} \quad (3)$$

If a point with deep color is mapped near the outer vertex of a pollutant axis, then such pollutant is the major air pollutant of that city. Figure 5 shows that smog is a national problem, because the AQIs of most of the cities are above 100. Furthermore, PM2.5 is the primary pollutant of the cities having a poorer air quality, due to the fact that most of the purple points (AQI > 300) are close to the PM2.5 axis.

In extreme cases, a point may be positioned outside of the Astrolabium. For example, assume the radius of the Astrolabium is 1, if the values of CO<sub>2</sub>, PM2.5 and NO<sub>2</sub> of a city are the maximum 1 and the other three attribute values are all 0, then the city’s coordinate is (2, 0). Therefore, we introduce a parameter  $k_i$  for each component coordinate, and Eq. 3 is changed as follows:

$$\begin{cases} (x, y) = \sum_{i=0}^n (u_i, v_i) \times k_i \\ k_i = M \times (\cos(|u_i, v_i| \times \frac{\pi}{2}) + 1) \quad (0 < M < 1) \end{cases} \quad (4)$$

where  $|u_i, v_i|$  represents the length of the  $i$ th component, ranging from 0 to 1, and  $k_i$  is monotonically decreasing. As  $|u_i, v_i|$  increases,  $k_i$  decreases more rapidly, and finally reaches  $M$  (when  $|u_i, v_i| = 1$ ). The user can manually adjust  $M$  to generate a desirable distribution of the small circles in the Astrolabium.

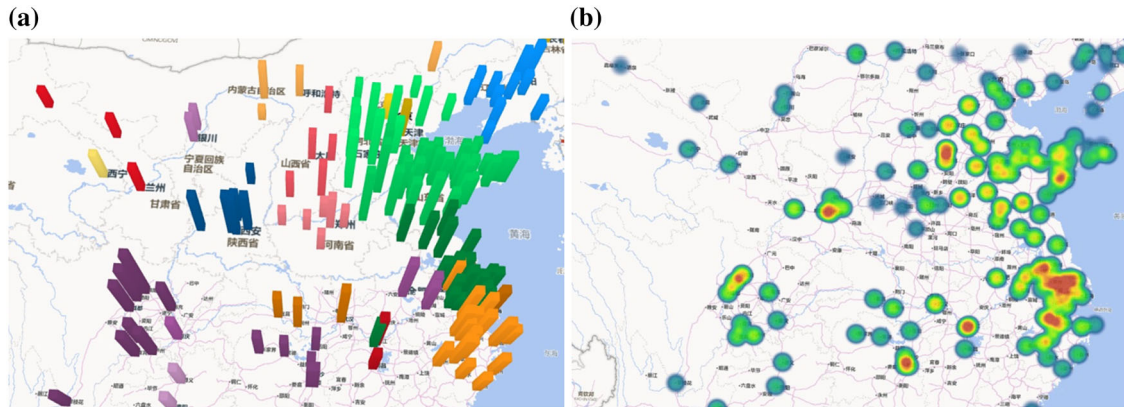
### 4.3.2 Air Component River

Since an Astrolabium can only show the component distribution at an instance of time, we design an Air Component River to represent the component variations of a city over a period. An Air Component River consists of multiple branches, each representing the content of a pollutant. As in Fig. 6, the width of the river indicates the sum of all the pollutants, and the width of each branch changes over time, visualizing the variation trend of each pollutant. Component Astrolabium and Air Component River can jointly show the trends and the air component distribution characteristics of different cities over time (see Fig. 11).

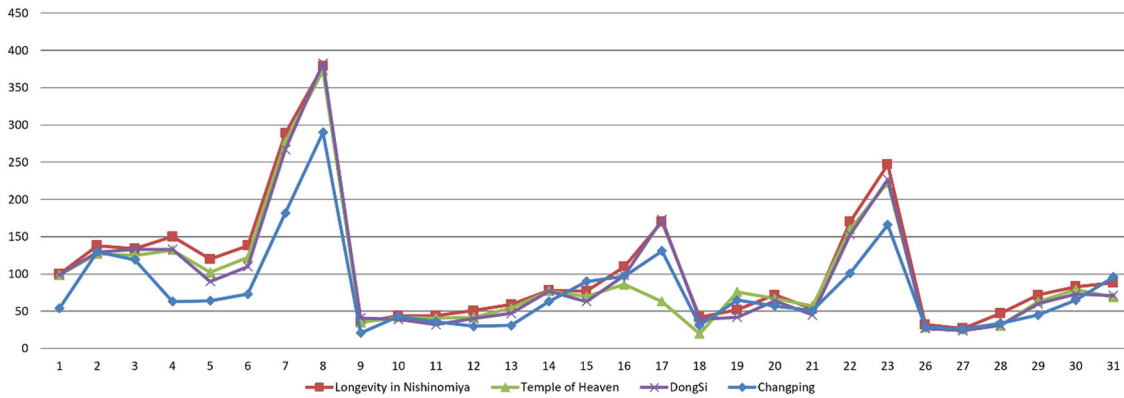
### 4.4 Aggregation View

We overlay bar charts on the map, making an Aggregation View for the interactive exploration and comparison of the air qualities in different cities.

The view in Fig. 7 is constructed based on a GIS platform. When using this view, the user first needs to select the pollutant, temporal interval and an aggregation function, then a 3D bar chart and a 2D heatmap are generated. These two plots are simply two different visual representations of the same data, and one of them can be easily selected according to the personal preference. Four aggregation functions (SUM, AVERAGE, MAX, MIN) used in DBMS are provided in this view. For example, if the user wishes to compare the max AQIs in a temporal interval of different cities, by selecting the MAX aggregation function, he/she obtains the view that reflects maximal AQI values in the two plots. This view also supports interactive operations often used in GIS platforms.



**Fig. 7** Aggregation View. **a** 3D bar chart. **b** 2D heatmap



**Fig. 8** Daily-average AQI of four stations in Beijing

## 5 Case study

This section reports the evaluation of our method using China Air Quality Observation Data and ECMWF re-Analysis data.

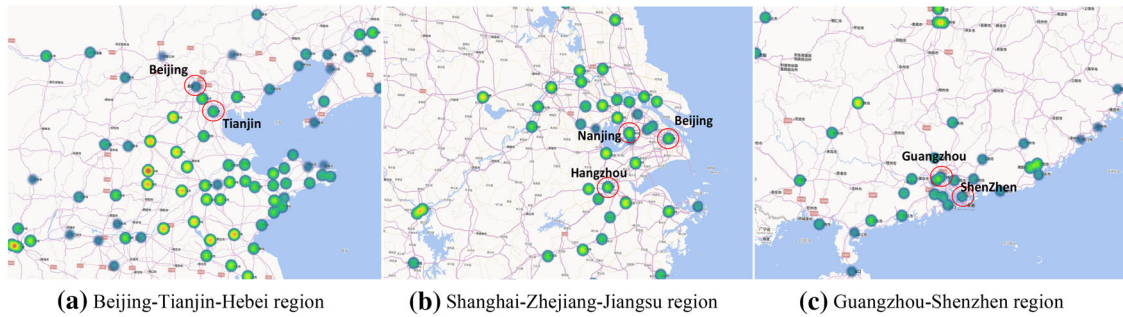
### 5.1 Data pre-processing

In a preliminary data analysis phase, we find the stations in one city have almost the same changing trend, as in Fig. 8. Therefore, we aggregate the data of all the stations in a city, and perform visual analysis using a city as the smallest unit. Because most cities are not exactly located in grid points, a linear interpolation algorithm is used to compute the meteorological attributes at the location of a city, weighted on the distances between the city and its four surrounded grid points.

We divide China into seven areas using the customary geographical division method in climate studies, and assign a color to each area. Each area consists of several provinces, each containing several cities. We thus construct a hierarchical structure to organize the geographic information. The ECMWF re-Analysis data are in NetCDF format. Through the NetCDF data operation interfaces of the oceanographic application platform, we conveniently obtain the meteorological attributes of the cities.

### 5.2 Overall distribution

We first analyze the overall smog distribution in China. Because December is considered to be the month when smog happened frequently, we download the smog data and meteorological data in December 2013.



**Fig. 9** The smog conditions in three important regions in China. *Red circles* indicate the capitals of the provinces in the regions. **a** Beijing, Tianjin and Hebei region. **b** Jiangsu, Zhejiang and Shanghai region. **c** Guangzhou and Shenzhen region

Using the example in Fig. 2a, we divided the cities into nine clusters based on the average AQI. We found that the numbers of cities in different clusters are almost identical and the average AQIs of all the clusters are higher than 50, implying that smog in China is a national phenomenon. By selecting the outmost ring that represents the cluster of the most serious smog, we find most of the cities in that cluster are in the west of Hebei province, and those cities are smogged during the entire month. East China is another area that suffered from smog, while the smog in Southwest, South and Northwest China are relatively moderate (see Fig. 9c). Inspecting the time series of all the cities in East China, we find the extreme smog occurred only at the beginning of the month.

The Aggregation View also proves our findings. Figure 9 visualizes the smog situations of three important areas in China. As the regional centers, Beijing, Shanghai, and Guangzhou all have high population densities and industrial production activities. Smogs in such cities are, however, not as serious as their surrounding cities. One reason may be that the small cities excessively aim to improve their GDPs and indiscriminately build new enterprises which may cause serious pollution. To the contrary, big cities have better urban planning and invest more funds on environment protection.

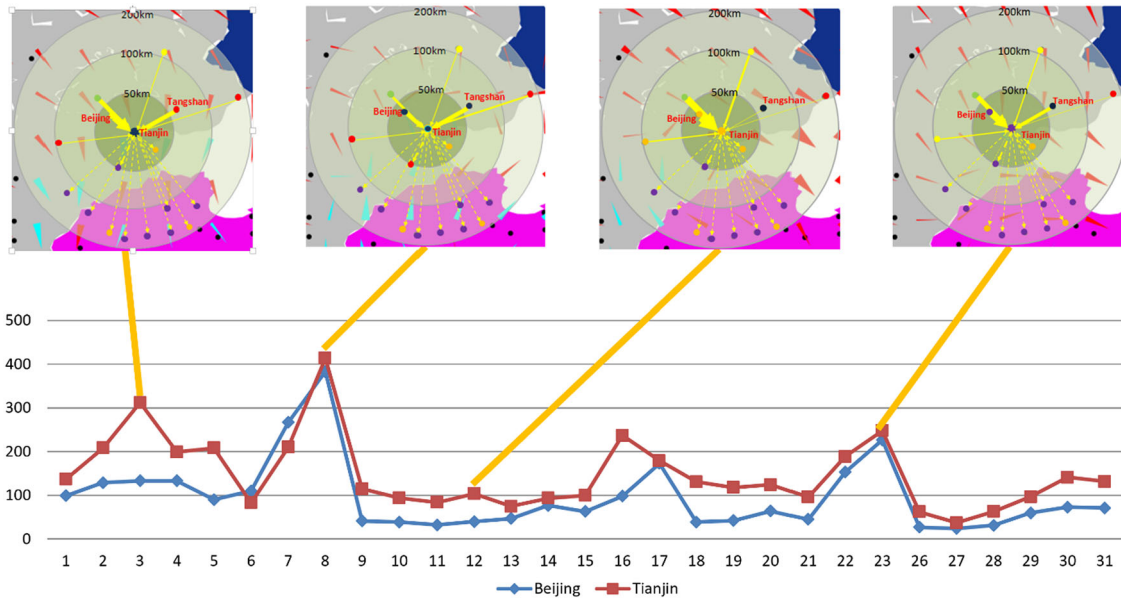
### 5.3 Correlation analysis

To reveal the correlation between meteorological attributes and smog phenomenon, we compare the air quality data of Beijing and Tianjin. These two cities are geographically close and have similar economic developments. As shown in Fig. 10e, the air qualities in the two cities are almost the same, but on 3 December, smog in Tianjin was very serious, while Beijing's air quality was normal. By querying the historical weather forecast, we found neither city experienced heavy rainfall or snowfall. Therefore, we tentatively hypothesized that the extreme smog in Tianjin came from other cities.

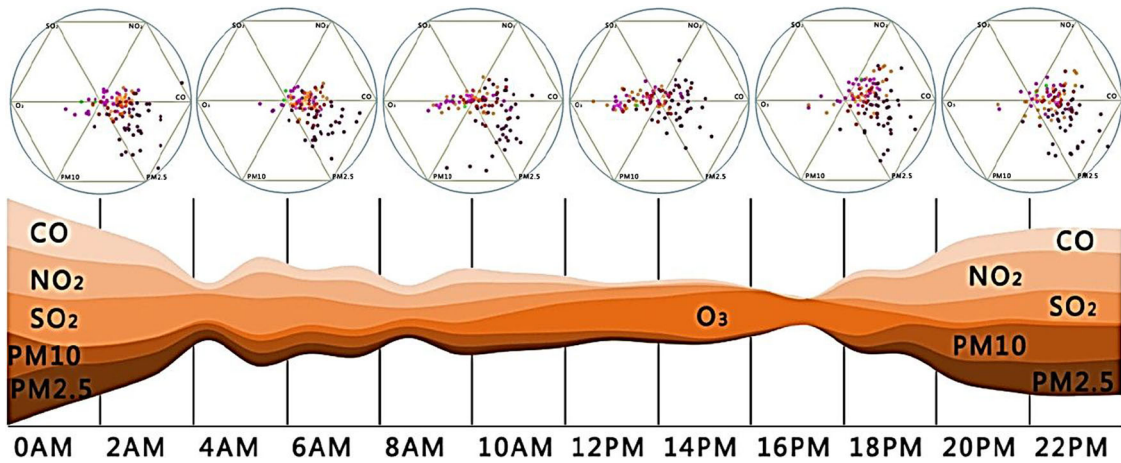
We first use the Correlation Analysis View to detect the smog sources of Tianjin. By visualizing the wind field, we found the winds in Beijing and Tangshan both blew to Tianjin (see Fig. 10a). Therefore, we suspected Tangshan as the pollution source of Tianjin's smog on that day. To prove our hypothesis, we continued to analyze the relationship between wind and smog on other days. We select the Tangshan as the target city, and view the influences of it on its neighboring cities. The radius of the observation range is set to 100KM to clearly view the air movement among such three cities (Beijing, Tianjin, and Tangshan).

We found when the arrowed lines from Tangshan to Tianjin and Beijing are dotted, which indicates the air movement is from Tangshan to other two cities, the air quality in Beijing and Tianjin are worse (see Fig. 10b, d). Instead, when the wind was east to west (the arrowed line from Tangshan to Tianjin are solid), Beijing and Tianjin both had relatively better air quality (see Fig. 10c). These findings are consistent with our hypothesis.

On 23th December, the wind was from Tangshan to Beijing and Tianjin (see Fig. 10d), but smog was not as serious as on 8th December. This may be related to inversion layer. On 8th December, there was an inversion layer around Beijing and Tianjin (see the highlighted points on the wind field plot), which hindered air movement and retained the smog. On the contrary, there was no inversion layer on 23th December. Smog was not so serious, although the wind direction was the same. Apart from the above cases, we also discovered many similar cases in other cities. The literature records that in 2008, many heavy



**Fig. 10** The correlation between AQI and wind and inversion layer



**Fig. 11** Air pollution contents change in 24 h

industrial enterprises were moved to Tangshan from Beijing in order to clear air for the Beijing Olympic Games, making Tangshan a current air pollution source of Beijing.

From these case studies, we learned extreme smog cases are not necessarily caused locally. It is highly related to the pollution levels of the surrounding cities, and the meteorological conditions.

#### 5.4 Air components change trend

Figure 11 shows the daily change of air components for all the 189 cities on 3rd December, 2013. Each Astrolabium represents the air components over 4 h. The point distribution in each Astrolabium is obviously close to the axis of PM2.5 and CO, implying that the problems of air pollution in most cities are associated with these two components. To view the component variation in a single day, we select Beijing as an example. We found that all the components except O<sub>3</sub> maintained a low level during the daytime, while the amount of O<sub>3</sub> increased obviously during the daytime. By consulting the related literatures, we know the

**Table 2** Questionnaire results

Factor	Highest	Lowest	Average	Variance
visual design	8	6	8.67	0.22
Interaction	9	7	8.20	0.48
Learnability	9	7	8.38	0.37
Performance	9	8	8.22	0.84
Functionality	9	5	7.61	1.21
Scalability	9	7	7.67	0.44

harmful gases, such as NO, NO<sub>2</sub>, SO<sub>2</sub>, CO and PM<sub>2.5</sub> will convert to O<sub>3</sub> under the influence of ultraviolet rays, which may cause the decrease of all the harmful gases except O<sub>3</sub>.

## 6 Evaluation

To evaluate the usability of our visualization approach, we have conducted an experiment with nine people volunteered as the experimental subjects. Four subjects are from National Ocean Technology Center with the education background of atmospheric science, while the other five are from the School of Computer Science and Technology, Tianjin University. The subjects are aged between 20 and 35, including 4 females and 5 males.

We first explained our approach and the prototype system to the subjects. After getting familiar with the system, they were asked to complete a questionnaire by providing scores between 0 and 10 on six aspects: visual design, interaction, learnability, performance, functionality, and scalability. We did not impose a time limit, yet every questionnaire was completed within 2 min. Table 2 shows the statistical results of the questionnaire.

The results show that the visual design scored the highest with the smallest variance, since all the subjects liked the design of our approach and believed that the four views were suitable for analyzing domain tasks. Interaction, learnability and the performance also received high scores, indicating that our approach is easy to learn and use. However, a subject pointed out that the four views could only satisfy specific tasks. This may account for the low scores of the functionality. In fact, our approach focuses on the primary tasks in air quality analysis, and we believe that no analysis tool can satisfy all kinds of the analysis tasks in environment science studies. Scalability also receives a relative low score. This may be because that the Global Distribution View might not be able to accommodate a great number of cities due to the small angle space of each sector. However, such effects are limited since the number of the cities having air quality observation data is less than 300. Furthermore, with several simple modifications of the metaphor, such as using a small circle on a clustering ring to represent the number of the cities in a region, etc., the Global Distribution View can accommodate a large number of cities.

An important limitation of our approach, pointed out by three subjects with atmosphere science background, is its correctness and accuracy. They thought our approach was too simple, lacking the complicated mathematical computing capability often used in traditional numeric models. Furthermore, although what we discovered have been verified, they remained doubtful about the correctness of our findings. In fact, all the existing models output results with inherent uncertainties, because of our limited understanding of many physical processes and the fact that not all the meteorological physics have been scientifically modeled. We hope that our approach capable of objectively visualizing actual observation data can be an effective alternative for experts to verify the conclusions generated by other methods. At the same time, our approach may also guide experts in more in-depth studies based on the patterns found in our approach, as a mechanism for hypotheses generation.

## 7 Conclusion

This paper has presented a comprehensive visualization approach to smog analysis. New visualization techniques have been integrated and used to analyze the smog problem in China using air quality data and ECMWF data. Having analyzed the smog variation patterns in Beijing, we consider our approach to be

effective and useful in real-world scenarios. However, due to the lack of fine-grained data and the participation of domain experts, our method can only be viewed as a qualitative analysis manner. Our approach could be improved in two aspects. First, visualization of the relationships between meteorological attributes and smog could be improved in a more intuitive manner. Second, cooperating with domain experts can make our approach more scientific and professional.

**Acknowledgments** The authors wish to thank Zhuang Cai, Meng-Yao Chen, and Chao Wen for their discussions. We are also grateful to the anonymous reviewers for their insightful comments that have helped us in improving the final presentation. The work was partially supported by Seed Foundation of Tianjin University (under Grant Number 2014XJJ-0003).

## References

- Brewer CA, Harrower MA (2009) ColorBrewer 2.0: color advice for cartography. Penn State University. <http://colorbrewer2.org/>
- Chen R, Zhao Z, Kan H (2013) Heavy smog and hospital visits in Beijing, China. *Am J Respir Crit Care Med* 188(9):1170–1171
- Chen J, Chen H, Zhang G, Pan JZ, Wu H, Zhang N (2014) Big smog meets web science: smog disaster analysis based on social media and device data on the web. In: Proceedings of the companion publication of the 23rd international conference on world wide web, pp 505–510
- Havre S, Hetzler B, Nowell L (2000) Themeriver: Visualizing theme changes over time. In: Proceedings of the IEEE Symposium on Information Visualization (InfoVis'00), pp 115–123
- Havre S, Hetzler E, Whitney P, Nowell L (2002) ThemeRiver: visualizing thematic changes in large document collections. *IEEE Trans Vis Comput Graph* 8(1):9–20
- Hulek R, Jarkovsky J, Boruvkova J, Kalina J, Gregor J., Sebkova K, Schwarz D, Klanova J, Dusek L (2013) Global monitoring plan of the Stockholm convention on persistent organic pollutants: visualization and on-line analysis of data from the monitoring reports. <http://www.pops-gmp.org/visualization/>
- Janicke H, Bottinger M, Mikolajewicz U, Mikolajewicz U (2009) Visual exploration of climate variability changes using wavelet analysis. *IEEE Trans Vis Comput Graph* 15(6):1375–1382
- Johansson S and Jern M (2007) GeoAnalytics visual inquiry and filtering tools in parallel coordinates plots. In: Proceedings of the 15th annual ACM international symposium on advances in geographic information systems
- Johansson J, Neset TS, Linner BO (2010) Evaluating climate visualization: an information visualization approach. In: Proceedings of the international conference on information visualization (IV'10), pp 156–161
- Landesberge TV, Bremm S, Andrienko N, Andrienko G, Tekusov M (2012) Visual analytics methods for categoric spatio-temporal data. In: Proceedings of the IEEE conference on visual analytics science and technology (VAST'12), pp 183–192
- Lee B, Riche NH, Karlson AK, Carpendale S (2010) Sparkclouds: visualizing trends in tag clouds. *IEEE Trans Vis Comput Graph* 16(6):1182–1189
- Li J, Meng ZP, Zhang K (2014) Visualization of oceanographic applications using a common data model. In: Proceedings of the 2014 ACM symposium on applied computing (SAC'14), pp 933–938
- Li J, Zhang K, Meng ZP (2014) Vismate: interactive visual analysis of station-based observation data on climate changes. In: Proceedings of the IEEE conference on visual analytics science and technology (VAST'14), pp 133–142
- Ma J, Xu X, Zhao C, Yan P (2012) A review of atmospheric chemistry research in China: photochemical smog, haze pollution, and gas–aerosol interactions. *Adv Atmos Sci* 29(5):1006–1026
- Malik A, Maciejewskiet R, Jang Y, Huang W (2012) A correlative analysis process in a visual analytics environment. In: Proceedings of the IEEE conference on visual analytics science and technology (VAST'12), pp 33–42
- Nguyen DQ, Schumann H (2010) Taggram: exploring geo-data on maps through a tag cloud-based visualization. In: Proceedings of the international conference on information visualization (IV'10), pp 322–328
- Nocke T, Schumann H, Bohm U (2004) Methods for the visualization of clustered climate data. *Comput Stat* 19(1):75–94
- Qu H, Chan WY, Xu A, Chung KL, Guo P (2007) Visual analysis of the air pollution problem in Hong Kong. *IEEE Trans Vis Comput Graph* 13(6):1408–1415
- Sharko J, Grinstein G, Marx KA (2008) Vectorized radviz and its application to multiple cluster datasets. *IEEE Trans Vis Comput Graph* 14(6):1427–1444
- Slocum TA, McMaster RB, Kessler FC, Howard HH (2009) Thematic cartography and geographic visualization, 3rd edn. Pearson Education, London
- Tominski C, Abello J, Schumann H (2004) Axes-based visualizations with radial layouts. In: Proceedings of the 2004 ACM symposium on applied computing (SAC'04), pp 1242–1247
- Wang C, Xiao Z, Liu Y, Xu Y, Zhou A, Zhang K (2013) SentiView: sentiment analysis and visualization for internet popular topics. *IEEE Trans Hum Mach Syst* 43(6):620–630
- Watson DF (1992) Contouring. Pergamon Press, New York
- Yannier N, Basdogan C, Tasiran S, Sen OL (2008) Using haptics to convey cause-and-effect relations in climate visualization. *IEEE Trans Vis Comput Graph* 1(2):130–141
- Zhou W, Cohan DS, Henderson BH (2013) Slower ozone production in Houston, Texas following emission reductions: evidence from Texas air quality studies in 2000 and 2006. *Atmos Chem Phys* 13(7):19085–19120