atmosphere

MDPI

Article

A Visualization Approach to Air Pollution Data Exploration—A Case Study of Air Quality Index ($PM_{2.5}$) in Beijing, China

Huan Li [1,2], Hong Fan [1,2,*] and Feiyue Mao [1,2,3,4]

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Luoyu Road 129, Wuhan 430079, China; lihuan@whu.edu.cn (H.L.); maofeiyue@whu.edu.cn (F.M.)
2 Collaborative Innovation Center of Geospatial Technology, Wuhan University, Luoyu Road 129, Wuhan 430079, China
3 School of Remote Sensing and Information Engineering, Wuhan University, Luoyu Road 129, Wuhan 430079, China
4 Hubei Collaborative Innovation Center for High-efficiency Utilization of Solar Energy, Wuhan 430068, China
* Correspondence: hfan3@whu.edu.cn; Tel.: +86-27-6877-8475

Academic Editor: Robert W. Talbot
Received: 24 November 2015; Accepted: 24 February 2016; Published: 29 February 2016

Abstract: In recent years, frequent occurrences of significant air pollution events in China have routinely caused panic and are a major topic of discussion by the public and air pollution experts in government and academia. Therefore, this study proposed an efficient visualization method to represent directly, quickly, and clearly the spatio-temporal information contained in air pollution data. Data quality check and cleansing during a preliminary visual analysis is presented in tabular form, heat matrix, or line chart, upon which hypotheses can be deduced. Further visualizations were designed to verify the hypotheses and obtain useful findings. This method was tested and validated in a year-long case study of the air quality index (AQI of $PM_{2.5}$) in Beijing, China. We found that $PM_{2.5}$, PM_{10}, and NO_2 may be emitted by the same sources, and strong winds may accelerate the spread of pollutants. The average concentration of $PM_{2.5}$ in Beijing was greater than the AQI value of 50 over the six-year study period. Furthermore, arable lands exhibited considerably higher concentrations of air pollutants than vegetation-covered areas. The findings of this study showed that our visualization method is intuitive and reliable through data quality checking and information sharing with multi-perspective air pollution graphs. This method allows the data to be easily understood by the public and inspire or aid further studies in other fields.

Keywords: air pollution; $PM_{2.5}$; visual exploration; spatio-temporal data analysis

1. Introduction

Since 2011, frequent occurrences of haze in China have become a cause for panic and routinely appear as a major topic in the media and on climate websites [1,2]. Thus, considerable research in various fields has focused on air pollution, attempting to solve the problem with different approaches [3–8]. Nevertheless, in many instances, the methods are not completely intuitive and are difficult for governmental officials and the general public to understand. Visual exploration [9–14] of air pollution with spatio-temporal data is a solution that makes complex data understandable because graphical representation is relatively intuitive.

Atmospheric particulate matter is a commonly used criterion to evaluate air quality [7,15,16]. The degree of adverse health effects depends on the size and composition of the particles [16]. $PM_{2.5}$ and

PM_{10} are defined as particles with diameters of 2.5 μm or less and 10 μm or less, respectively; these parameters are usually measured using air quality index (AQI). AQIs are calculated from the particle concentration at monitoring stations expressed as micrograms per cubic meter [17]. According to the technical regulation on ambient air quality index (on trial) [18], the air pollution index for $PM_{2.5}$ is divided into six levels, namely, 0–50, 51–100, 101–150, 151–200, 201–300, and greater than 300. With these levels, we can identify the severity of air pollution.

Literature on chemical and remote-sensing fields greatly contributes to air pollution analysis. Examining air composition with related chemometric techniques can achieve effective analysis of pollutant sources [19], which is suitable for micro and local level analyses but is not satisfactory for spatio-temporal pattern exploration of larger areas. Satellite remote-sensing assessment, subsequent mapping of the spatial distribution of aerosols [4,20], and air pollution data measured by satellite-borne sensors or ground-based monitoring stations [21,22] identified a strong relationship between $PM_{2.5}$ and PM_{10}. However, complicated calibration and data processing are executed with professional software, most of which are standalone and specialized and thus cannot be easily understood or shared with the general public.

By dealing with graphical representations of complex data, conventional plots, such as scatter plot, are used to analyze time-series data and show the correlation among various factors in air pollution exploration [7,23]. These plots are simple and have been widely used for a long time, but they cannot effectively express spatial relationships and are less attractive than the newer multi-perspective visualizations and graphs. Moreover, the rapid development of web technology allows interactive visualizations to combine various technologies [24], such as HTML for page content, CSS for aesthetics, JavaScript for interaction, and SVG for vector graphics. These technologies render sharing intuitive information highly convenient.

Many spatial distribution explorations and complex representations of air pollution [3,8,25] rely on standalone or proprietary software products like ArcGIS. VIS-STAMP [26] is a software that provides tools for users to generate self-organizing maps, parallel coordinate maps, map matrices, and reorderable matrices (a type of heat map). This software allows users to visualize multivariate statistical analysis intuitively, as well as explore and understand spatio-temporal and multivariate patterns. Nevertheless, this type of exploration does not consider the extendable and sharable functions that are important for the public to acquire multi-perspectives and latest information on air pollution.

Numerous researchers focus on systematic theories of visual exploration for spatio-temporal data. Kraak [12] claims that graphics can reveal patterns that are not necessarily visible when conventional map display methods are applied, demonstrating the usefulness of geovisualization. The idea of a generic scenario extends the idea of geovisualization following the schema developed [9,13]. A general framework was proposed then for using aggregation in visual exploration of massive movement data [27]. Similarly, another framework [28] was discussed for spatio-temporal visualization with the visual method space–time cube. However, these frameworks must be extended further and adapted for the visual exploration of air pollution data considering data quality, correlations in multivariate data, and spatio-temporal pattern analysis.

Air composition analysis and remote-sensing methods require complex computations and are time consuming. Moreover, the current graphical approaches for air pollution analysis lack interactive and sharable multi-perspective visualizations. The public urgently needs rapid and reliable knowledge on air pollution to make daily decisions. Similarly, most of the government staff are not professionals and need intuitive understanding of the conditions before they can execute any actions against the increasingly serious air pollutants. Thus, a visual methodology is needed for efficient and reliable exploration, particularly in the case of air pollution data, to improve the depth, readability, and accuracy of data analysis. We were motivated by the idea of visual thinking and the geovisualization concept [12], which may be well suited for the exploration of air pollution data. Building on the extensive body of previous work, our paper presents a visualization exploration method that realizes the process of observation–hypothesis–verification. This method was tested and validated in a year-long case study

of the air quality index (AQI of $PM_{2.5}$) in Beijing, China. The useful findings of air pollution in Beijing show that our new method extends the existing work and fills a gap in the research, focusing on visual exploration to support various applications, such as knowledge-based decision making and aided research of air pollution.

2. Method and Data

Workflow for visualizing air pollution data is shown in Figure 1 and consists of five blocks, namely, Database, Preliminary analysis, Hypotheses, Verification, and Application. Our previous work, a Real-time Sensor Data Provision system (ReSDaP) [29–31], achieved the on-demand data provision for the Database. Data are then checked and analyzed by the Preliminary analysis block through various basic visual methods, such as bar charts, heat tables, and line charts. Based on these data, hypotheses can be built and verified through further visual analytics. Finally, the findings of previous steps can be used for applications.

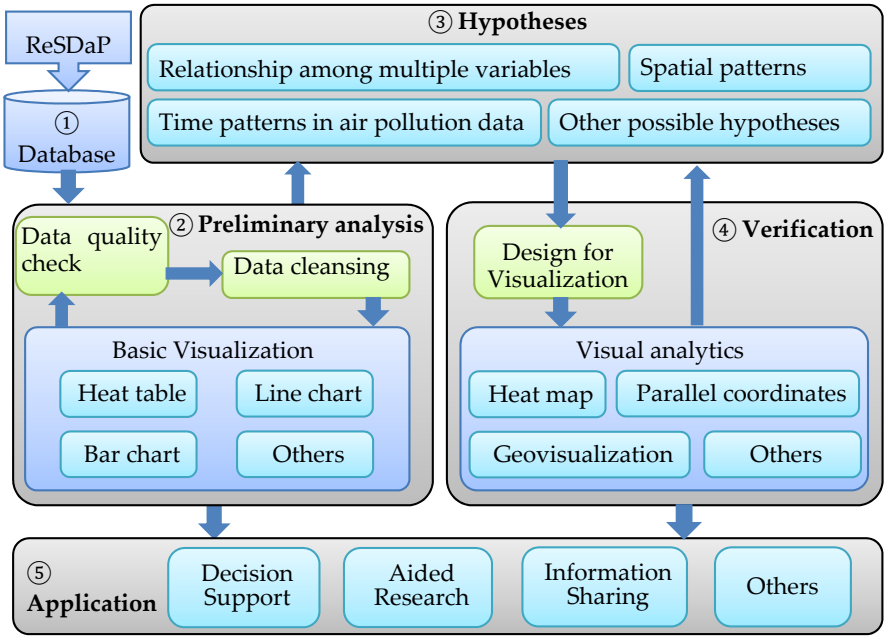


Figure 1. Visualization workflow for air pollution data.

Data accessing and storage are displayed as Step (1) in Figure 1. The backend data are acquired from our previous work through the real-time data provision system architecture for sensor webs (ReSDaP) [29]. Air pollution data often contain time and geographic location and can include an AQI, a $PM_{2.5}$ concentration measurement. The data may also include other information related to air pollution, such as weather and economic development data. These data exist in various formats and are stored based on these formats in different types of databases. Data similar to tab-separated values, comma-separated values, spreadsheets, and other types of data tables can be directly stored in a relational database, such as the open source database MySQL. For unstructured data, such as text, pictures, and videos, the NoSQL (not only SQL) database can be used for storage, such as the MongoDB open source database that can also store structured data. Through ReSDaP [29], air pollution data are directly stored in the database, providing support for data access and storage.

The Preliminary analysis block is shown as Step (2) in Figure 1. The main goal of this layer is to carry out an overall analysis of the data, perform basic data cleansing, and obtain several elementary discoveries. The next block shows the Hypotheses as Step (3). Verification is shown in Step (4), which is used to verify previously tested hypotheses. The main processes in this block include visualization design, data preprocessing, and visual analytics to achieve visualization flexibility with various

methods, such as heat maps, parallel coordinate plots, heat circles, and calendar views. The last step is the application block seen as Step (5) in Figure 1. The following subsections will describe these four blocks in detail.

2.1. Preliminary Analysis

The Preliminary analysis block allows users to perform preliminary global visualization, explore the data to examine possible inferences, and provide clues and guides for the next block, Hypotheses. This block enables global analysis to verify the correctness and the completeness of the data, which permits preliminary data visualization with basic graphs, such as heat tables, line charts, and bar graphs, for a simple and rough overview of the data. If necessary, this step can be repeated to perform new data inspection and processing. A reliable process for data quality checking and cleaning is needed before information extraction from the original data. An inspection of the raw data reveals defects; thus, more reliable results can be acquired after visual support. Rearranging the rows and columns with drag-and-drop and partially selecting some data items through check boxes are highly convenient, which make finding missing data and logical errors in the data easier.

The Preliminary analysis block is rapid and easy with visualization tools. The complicated algorithms for accessing data, chemical analysis [19] of air, and remote-sensing [4,20] methods require proprietary software and knowledge to interpret the data or are time consuming and difficult to understand. The same is true regarding the analysis of the spatio-temporal distribution of air pollution values in space and time [3,7,23,25,26,32]. An Online interaction user interface can perform rapid basic visualizations to find preliminary results and inspect data defects with its flexible data formatting and special functions. Many visualization tools are free, open source, easy to acquire, and extendable. JavaScript is one of the techniques that make it convenient for the public to access and understand the data.

2.2. Hypotheses

Basing on basic visualization and data cleansing, possible patterns in air pollution data can be hypothesized, such as relationships among multiple variables, time-related fluctuation of pollutant concentrations, and spatial distributions. This part is based on adequate basic visualization to find specific characteristics or trends shown in the charts. After proposing the hypotheses, verification is needed as described in the succeeding section.

2.3. Verification of the Hypotheses

Specific visualization methods were designed for testing and verification of possible inferences. To verify the hypotheses by visual analytics, a visualization design will help users determine the next steps for further analysis. The Visual analytics with more refined visualization and analysis can be used to verify the thoughts in the Hypotheses block. If hot spots were found in this part, more hypotheses could be built and further verification would be needed. Thus, this can be a recursive process.

We achieved a better visual effect from multi-perspective visualization for spatio-temporal analysis. Conventional plots, such as scatter and line plots, are usually used for the time-series data exploration of air pollution [7] without spatial analysis and do not include the new attractive graphs. We analyzed the characteristics of various types of graphs and found that the circular heat map is extremely suitable to display time series data and calendar view [33] is useful for displaying years of daily data. This method provides a new perspective on time series data and a means to comprehensively understand the conventional line charts or matrix plots that fail to deliver because they separate the data into several parts according to seasons or years. We also used the geovisualization graphs to represent the spatial distribution of air pollution data clearly.

2.4. Application Layer

Visualization highlights the regularities in the data to be verified by later statistical analysis. Both basic visualization and visual analytics lead to discovery and provide support for applications. In the Application layer, relevant findings are summarized from the visual analytics of the Preliminary analysis and Verification blocks to the appropriate treatment programs. For example, if long-term positive correlations were found between carbon dioxide concentrations and temperatures all over the world, controlling the irregular emission of carbon dioxide may be an approach to address global warming. Thus, the visual display of data supports decision-making by experts and non-experts alike. However, anomalies seen in the charts that illustrate data hot spots and allow users to identify data for further statistical analysis. At the same time, data visualization can act as aided research of other methods in the air pollution field, and sharing interactive findings through the web. Therefore, the application block in the workflow can be coordinated with other systems, providing a basis for a comprehensive model of a phenomenon.

We can further stress the advantages of our method by comparing them with previous methods for visual analytics. The simple model of visualization for visual analytics [9] was too abstract and coarse to be used for the analysis of air pollution data with graphs. Andrienko and Andrienko [27] proposed a general framework using aggregation in visual exploration of massive movement data. A similar framework for spatio-temporal visualization was suitable for movement data with the visual method space–time cube [28]. It was a problem-driven process with domain experts as users. Nevertheless, our method contains a data-driven process with multi-perspective spatio-temporal visualization that is easily understood by expert and non-expert users alike. Moreover, we described an efficient visual method to check data quality and enable our method to acquire more reliable results.

Visualization value can be assessed from three viewpoints: technology, art, and empiric science [9]. Our method is efficient and extendable from the technological viewpoint and beautiful from the artistic viewpoint. Meanwhile, it follows the "observation–hypothesis–verification" process of the scientific method and thus complies with the demands of empirical science. In Section 3, this method will be tested and validated in a year-long case study of the AQI of $PM_{2.5}$ in Beijing, China.

2.5. Experimental Data

Two datasets are utilized for the case study of air quality in Beijing, the capital of China. The first dataset is historical hourly $PM_{2.5}$ data (2009–2014) of Beijing obtained from the US Department of State air quality files available on its website [34] as measured at the US Embassy in Beijing. Observation values include $PM_{2.5}$ concentration, with parameter concentration units in micrograms per cubic meter ($\mu g/m^3$) transformed to AQI values for our case study. However, these data are not completely verified or validated as indicated in the data statement. These data were used to demonstrate a practical application of the visualization approach, as well as its feasibility and efficiency. The data are referred to as the USE data in the rest of this paper.

The second dataset was created from the U-Air project [35] and comprises one year (8 February 2013 to 8 February 2014) of air quality data from 36 air quality monitoring stations in Beijing, all with geographic coordinates. The stations are shown in Figure 2. The observation data include time, AQI of $PM_{2.5}$, PM_{10}, NO_2, temperature, humidity, wind speed, and weather. Notably, the original weather data have several anomalies; thus, we decided to explore the relationship between pollution and the rainy weather with precipitation data only [36]. These data are termed the U-Air data in the rest of this paper.

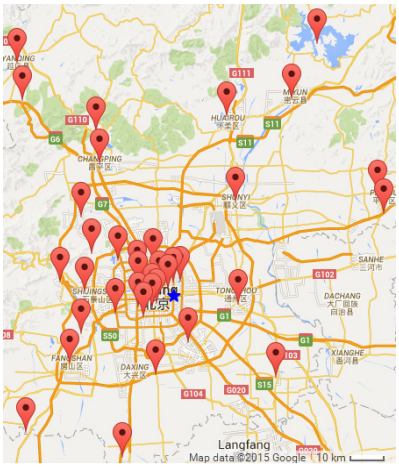


Figure 2. Distribution of Beijing observation stations. Blue star represents the US Embassy (39.92 N, 116.44 E) located in the city center.

3. Results and Discussion

3.1. General Analysis and Hypotheses

Before original data records could be used to convey information, preliminary analyses are necessary to evaluate data quality and find general facts in the data itself. Moreover, data-driven hypotheses could be built to guide further analysis. In this section, we will cover these steps without going into details. Readers can find more details in the Supplementary Materials as the main goal of this paper is the flexible visual exploration method for air pollution. Missing data and logic errors may occur in the original data. Therefore, rapid data inspection is needed before they can be used to acquire reliable findings. Visual analytics is an intuitive and clear method for this purpose. Data misreads can be avoided to some extent if these defects are found and processed through graphs. We will show how the data quality of the original U-Air data can be verified and validated with visualization methods.

With tables or scatter plots, missing data can be easily found. As presented in Figure 3, stations 23–36 from March to October have no observations, which clearly show the missing blocks to be handled. Moreover, logic errors could be found with this visualization type. For example, when setting rows as months and columns as weather, the evident relationship between weather and seasons could be observed intuitively. Beijing has a sub-humid warm temperate continental monsoon climate [37] on the northern hemisphere; thus, it is unlikely to snow in summer.

These problems can be handled in several ways. Multi-source data fusion may supplement the missing data and replace the part with logic errors. In this case study, the precipitation data were integrated to determine whether rainy weather can affect pollutant concentration. Interpolation for missing or coarse records is another commonly used method [23]. In addition, data could be divided into parts and analyzed separately to be used for the missing data problem in this case study.

Figure 4 shows the missing data in stations 23–36 from March to October. To minimize the effect of these missing data in the analysis, we separated the data into two parts and outlined them in Table 1. The first strategy involves the time pattern analysis of $PM_{2.5}$ for stations 01–22 for all months. This analysis includes visualizing the average AQI of $PM_{2.5}$ by month, day, and hour. The second strategy is spatial distribution analysis of $PM_{2.5}$ for observations in January, February, November, and December because all stations in these months have complete data records. With this separated analysis, more reliable exploration can be performed.

station_id \ month name	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Totals
001021	140.59	159.37	209.10	124.50	157.75	183.03	159.58	116.51	150.75	168.78	109.15	106.74	**144.39**
001022	137.55	169.40	224.98	121.39	148.32	164.07	132.47	124.73	156.80	163.42	99.74	106.85	**143.28**
001023	118.50	103.25									77.40	72.49	**95.73**
001024	110.06	94.03									67.28	84.59	**92.84**
001025	114.26	98.59									66.85	91.87	**95.89**
001026	137.74	112.25									66.66	97.94	**109.79**
001027	110.72	103.19									64.42	69.88	**89.32**
001028	102.45	97.86									73.66	57.92	**83.55**
001029	110.13	107.84									62.80	73.03	**90.26**
001030	91.29	94.32									52.28	65.93	**76.62**
001031	79.59	87.38									49.75	50.11	**66.27**
001032	77.54	82.62									51.45	35.69	**60.82**
001033	108.91	93.95									63.45	83.24	**91.36**
001034	186.91	94.24									117.17	184.96	**163.53**
001035	189.13	90.99									108.21	176.93	**159.36**
001036	192.41	116.25									147.03	188.07	**174.60**
Totals	**124.89**	**114.20**	**217.04**	**122.95**	**153.10**	**172.13**	**144.52**	**120.65**	**153.78**	**166.10**	**79.96**	**96.74**	**115.75**

Data missing

Figure 3. Missing data identified through an organized table.

year	month name \ day name	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Totals
2013	Feb	180.16	268.69	183.10	176.82	114.25	239.01	302.13	**205.73**
	Mar	58.03	179.38	276.28	332.46	219.89	160.09	305.70	**217.48**
	Apr	124.76	133.01	109.77	85.10	118.51	120.16	146.94	**118.33**
	May	166.60	161.29	125.02	151.17	156.05	144.32	143.52	**149.03**
	Jun	159.83	160.57	176.73	174.61	212.15	171.31	217.92	**182.84**
	Jul	151.83	95.96	97.78	124.66	143.39	161.24	146.09	**129.80**
	Aug	115.02	161.33	147.08	111.15	119.79	97.17	118.01	**125.13**
	Sep	143.59	118.45	127.80	149.10	166.40	161.10	152.65	**145.76**
	Oct	162.45	120.53	143.44	157.64	186.31	167.99	178.59	**157.06**
	Nov	51.77	99.65	162.63	93.67	196.46	45.49	26.42	**97.47**
	Dec	148.88	134.56	92.29	42.48	75.29	104.66	150.25	**110.59**
2014	Jan	112.62	102.25	164.26	186.46	136.50	142.25	102.86	**136.58**
	Feb	8.13	75.07	144.21	139.86	122.75	130.06	43.06	**97.78**
	Totals	**132.21**	**131.46**	**138.07**	**138.08**	**146.23**	**140.26**	**147.88**	**139.10**

Figure 4. Heat table of month and week mean air quality index of $PM_{2.5}$.

Table 1. Visualized design for discovery of defects in the U-Air data.

Stations (ID)	Months	Mean AQI of $PM_{2.5}$
Part (01–22)	All months	Relationship with Month, Day, Hour
All (01–36)	Part (1, 2, 11, 12)	Relationship with Geo-location

We performed the global visualization of data using basic charts. Based on these visualizations, possible hypotheses were generated for the visualization design program. We used the U-Air data. According to Table 1, the first strategy is used to determine whether a temporal pattern is evident in the data from stations 01–22 for all months. Through basic visualization, as shown in the heat table in Figure 4, higher $PM_{2.5}$ concentrations are observed from Friday to Sunday from February to March than during other times in this period.

The line chart shown in Figure 5 indicates that the highest mean AQI for $PM_{2.5}$ was in March, whereas the minimum was in November.

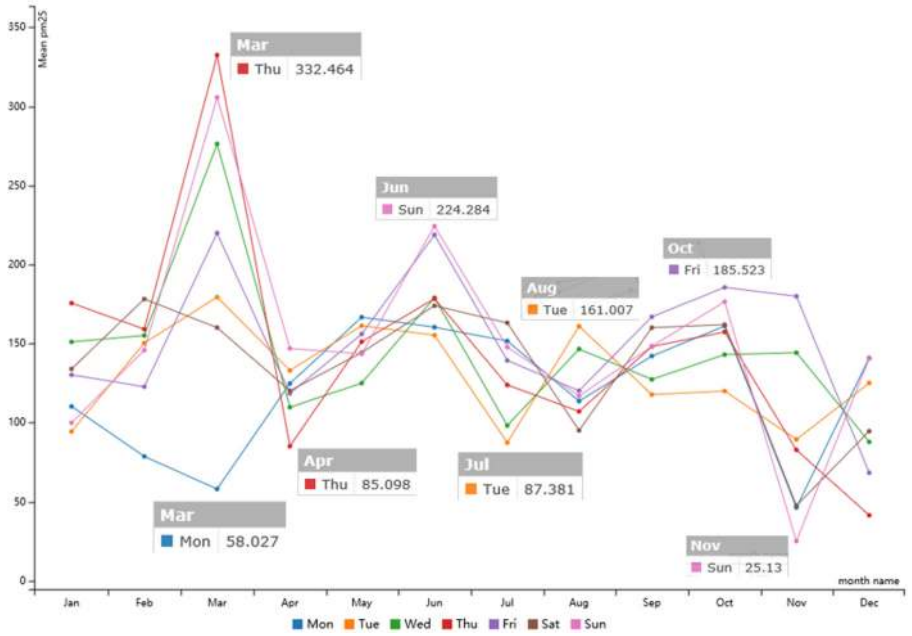


Figure 5. Monthly mean values of $PM_{2.5}$. Colored lines represent days from Monday to Sunday.

AQI variation in March is large because heating is sometimes provided. However, heating can be unnecessary because the weather is getting warmer. The mean AQI for $PM_{2.5}$ was relatively high during January, June, and October.

As shown in Figure 6, lines in the area chart display the same trend over four months, indicating that contaminants had spatial distribution characteristics. Stations 01, 03–04, 06–16, and 20–22 had high values; stations 05, 17–19, and 34–36 exhibited extremely high pollutant concentrations; and Stations 02 and 23–33 showed the lowest values among all the monitoring sites. These results demonstrate that air pollutants have a strong spatial distribution.

A heat matrix is a representation of the correlations between the data with a color matrix. The correlation value is a Pearson product moment correlation coefficient, r (Pearson's r for short, with the range of -1 to 1), which is calculated using Equation (1) given a series of n measurements of variables X and Y written as x_i and y_i, where $i = 1, 2, \ldots, n$. This computation yields a correlation matrix in which each i, j element is equal to the r value between the X and Y variables.

$$r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{\left[n\sum x_i^2 - (\sum x_i)^2\right]\left[n\sum y_i^2 - (\sum y_i)^2\right]}} \quad (1)$$

A heat matrix can be used to intuitively discover the overall relevance of data; thus, it can be used for preliminary analysis of the data. As shown in Figure 7, colors represent Pearson's r, and color ranges from red to blue by gradient. The area marked with red dotted boxes in Figure 7 indicates several primary results. The AQIs of $PM_{2.5}$, PM_{10}, and NO_2 showed strong positive correlations, whereas the wind had a negative correlation with the AQIs of $PM_{2.5}$, PM_{10}, NO_2, and humidity. When the AQI of $PM_{2.5}$ is high, high AQI of PM_{10} and NO_2 are obtained, and the wind speed may be low. Through the general analysis, we constructed three hypotheses and the corresponding visualization design for the air pollution exploration. The first hypothesis is that a correlation exists among the AQI of $PM_{2.5}$, PM_{10}, and NO_2, and wind speed. This correlation can be visualized through scatter plots that show the relationship between any two factors. The second hypothesis is that a regular time pattern exists for air pollutants. Heat maps would be an ideal visualization method, including a circle heat map and a calendar view. The third hypothesis is that air contaminants possess a geographical

distribution. This distribution can be represented appropriately by geovisualization, a method suitable for illustrating continuous spatial distributions.

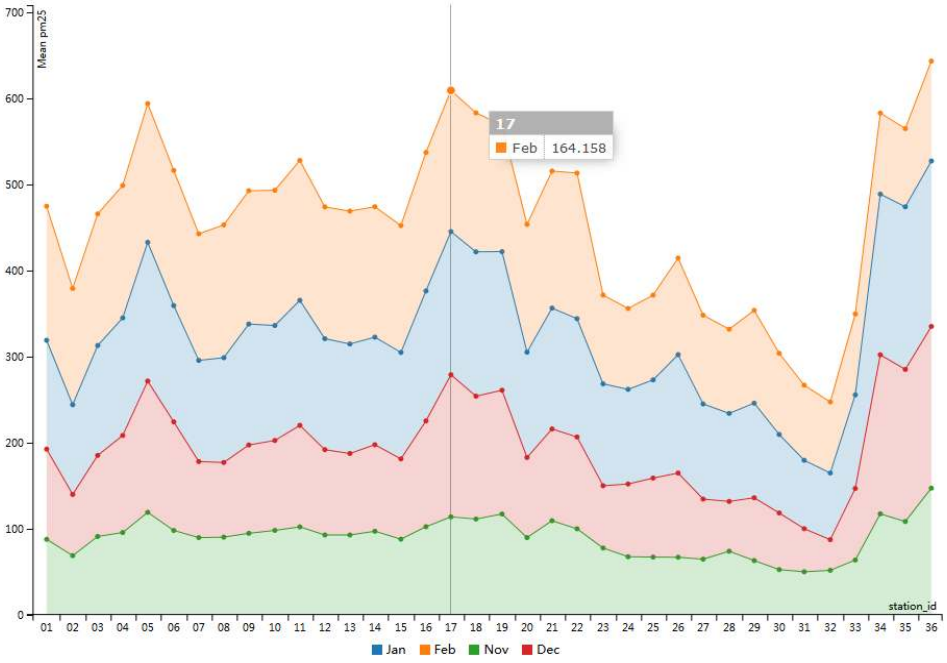


Figure 6. Area chart displaying the mean AQI of $PM_{2.5}$ on each observation site (id: 01–36) in January, February, November, and December.

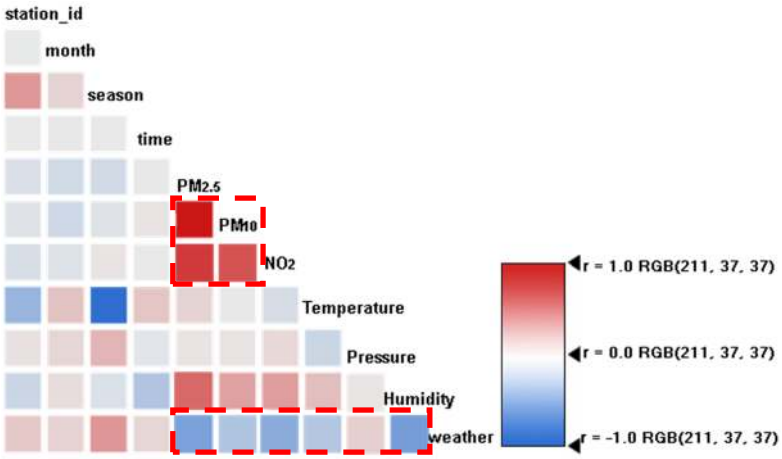


Figure 7. Correlation matrix of the U-Air data.

3.2. *Multi-Perspective and Various Visual Analysis*

We explore results in this section using multi-perspective analysis and various visual graphs. The relationship between pollutants and weather factors is analyzed first, followed by the presentation of temporal characteristics of $PM_{2.5}$. Afterward, spatial features of air pollution are shown and studied with stations overlaid on a satellite map. The last subsection includes the findings of air pollution in Beijing.

3.2.1. Relationship between Multiple Factors

Section 2.2 introduces the hypothesis that a relationship exists among pollutants and wind speed. In this section, plots are used to visualize and explore the relationships between the pollutants and weather using the U-Air data. Figure 8 shows that the three pollutants are strongly positively correlated because the Pearson's *r* values are rather high. However, wind speed has a negative correlation with $PM_{2.5}$, PM_{10}, and NO_2, which have high negative *r* values. Precipitation exhibits an extremely weak negative correlation with the pollutants. The strong positive correlation between $PM_{2.5}$ and PM_{10} shown in Figure 8 verifies the findings in [7]. Because $PM_{2.5}$ and PM_{10} are defined as particles with diameters of 2.5 μm or less and 10 μm or less, respectively, the relationship reveals that the particle density distributed stably according to their sizes. The positive correlation between pollutants indicates that $PM_{2.5}$, PM_{10}, and NO_2 may be emitted by the same sources, or one may be emitted by the transformation of another through some type of chemical mechanism [38]. To determine the specific reasons, a combined physical and chemical analysis of pollutants is desirable [19].

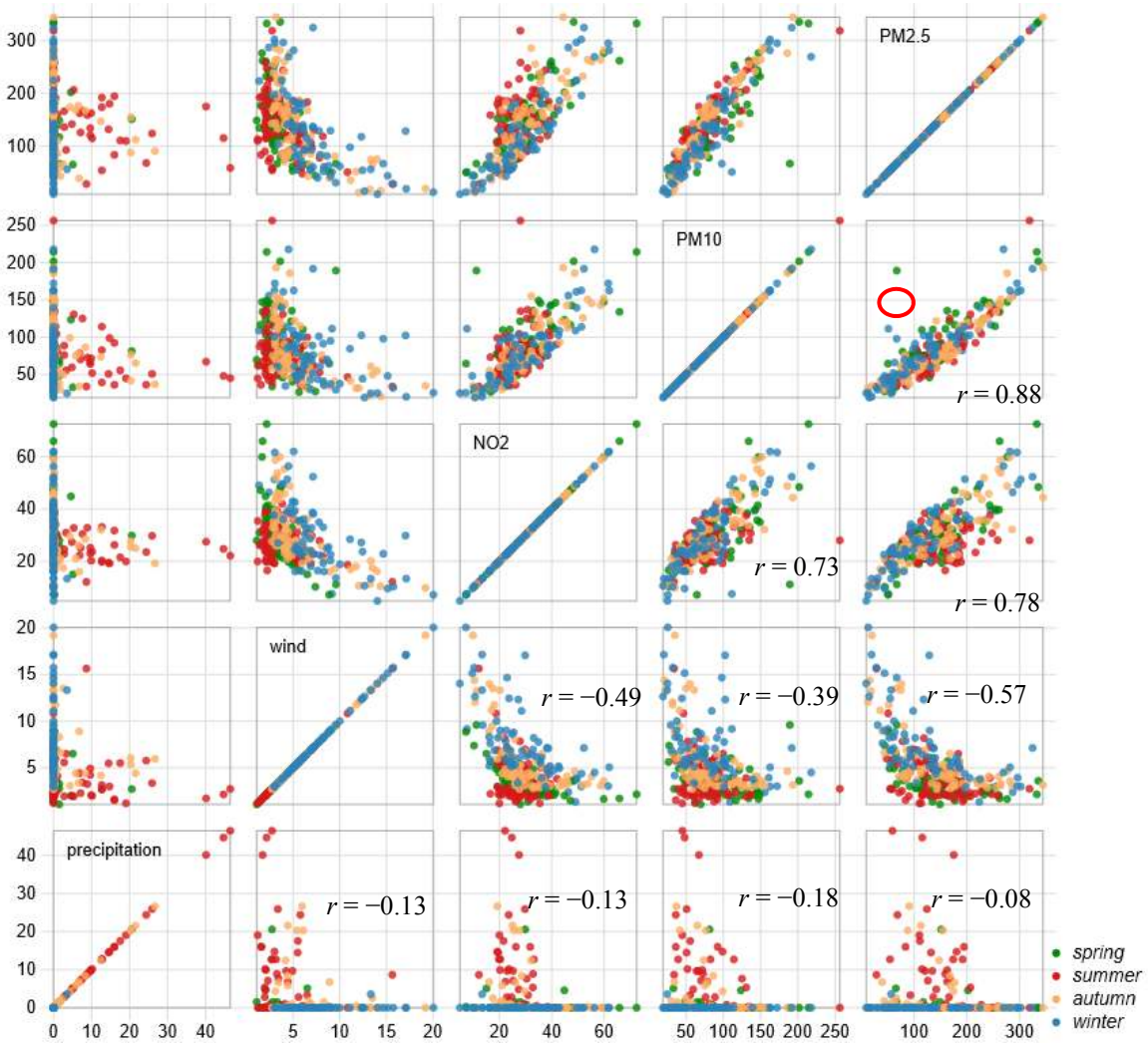


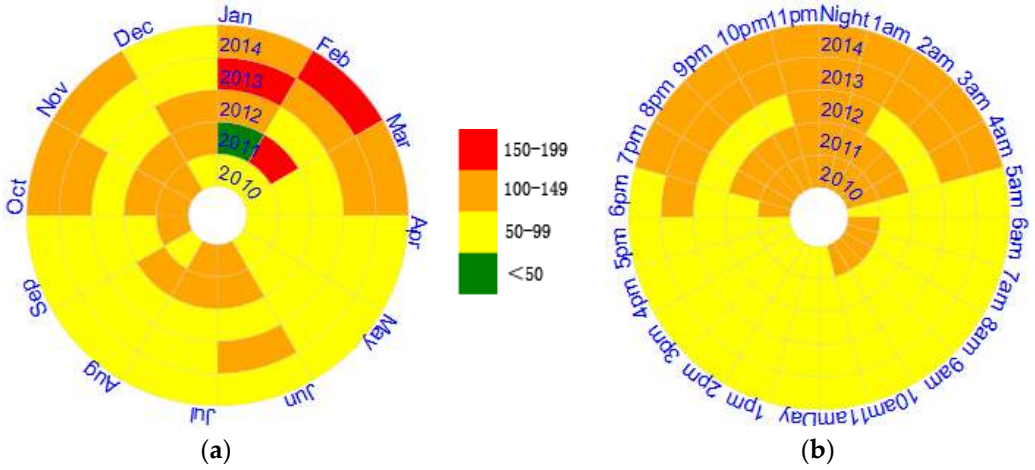
Figure 8. Scatter plots showing relationships among $PM_{2.5}$, PM_{10}, NO_2, wind, and precipitation. Different colors represent four seasons, where *r* is the correlation coefficient of two factors. Red circle marks a special point that is markedly different from the others.

The red circle marked in Figure 8 shows a different record in spring, which is far from the main cluster. This data item was found on 9 March 2013. The AQI of PM_{10} exhibited an extremely high

value of 189.28, whereas that of $PM_{2.5}$ had a low value of 66.95. At that time, Beijing was experiencing a sandstorm caused by the northerly wind. Moreover, a sandstorm contains particles with large sizes that may lead to high concentrations of PM_{10} despite the consistency of $PM_{2.5}$ [39]. This point is different from that of the others. Figure 8 also shows that pollutants and wind speed are negatively correlated; that is, larger wind speed results in lower pollutant concentration. Strong winds will accelerate the spread of pollutants, and a mixture of fresh air may decrease pollutant concentration. Moreover, precipitation has a weak negative correlation with the pollutants. Precipitation can wash the $PM_{2.5}$, which can be replenished very soon in Beijing. Summer and autumn have strong precipitation but there is extremely light rainfall in spring and autumn; this conforms to the sub-humid, warm, temperate, and continental monsoon climate in Beijing [37].

3.2.2. Temporal Characteristics

The USE data, which had long-term observations, were used in this section to determine the temporal characteristics of the mean AQI values of $PM_{2.5}$. The average AQI of $PM_{2.5}$ in five years is visualized through two circular heat maps; one map is a monthly variation and the other is an hourly variation. A daily concentration of $PM_{2.5}$ is presented as a calendar map, which is useful for an intuitive inspection of the severity of pollutants. As shown in Figure 9, pollutant concentrations are extremely high from 2010 to 2014. Figure 9a shows a circular heat map with average AQI for every month in selected years, whereas Figure 9b represents 24-h periods for selected years. Basing on the charts, AQIs of not less than 150 apparently occurred in January and February, whereas AQIs of more than 100 occurred in September, October, January, and February. Moreover, the AQIs of more than 100 were concentrated from 6:00 p.m. to 4:00 a.m. during the selected years.



(**a**) (**b**)

Figure 9. Circular heat maps from 2010 to 2014 are selected: (**a**) year and month circular heat map showing the monthly air pollution concentration; and (**b**) year and hour circular heat map showing the daily air pollution concentration.

As shown in the calendar view (Figure 10), days with a mean AQI of $PM_{2.5}$ greater than 200 (heavily polluted) were concentrated from January to March and October to December. No significant difference was observed in $PM_{2.5}$ between weekdays and weekends, as shown in Figure 10. These graphs of temporal characteristic show that the average concentration of $PM_{2.5}$ in Beijing was greater than the AQI value of 50 over the six-year study period; therefore, the residents suffer from long-term moderate pollution. The concentrations are more serious from October to March and from 6:00 p.m. to 4:00 a.m. on average. People are advised to stay indoors during those time periods.

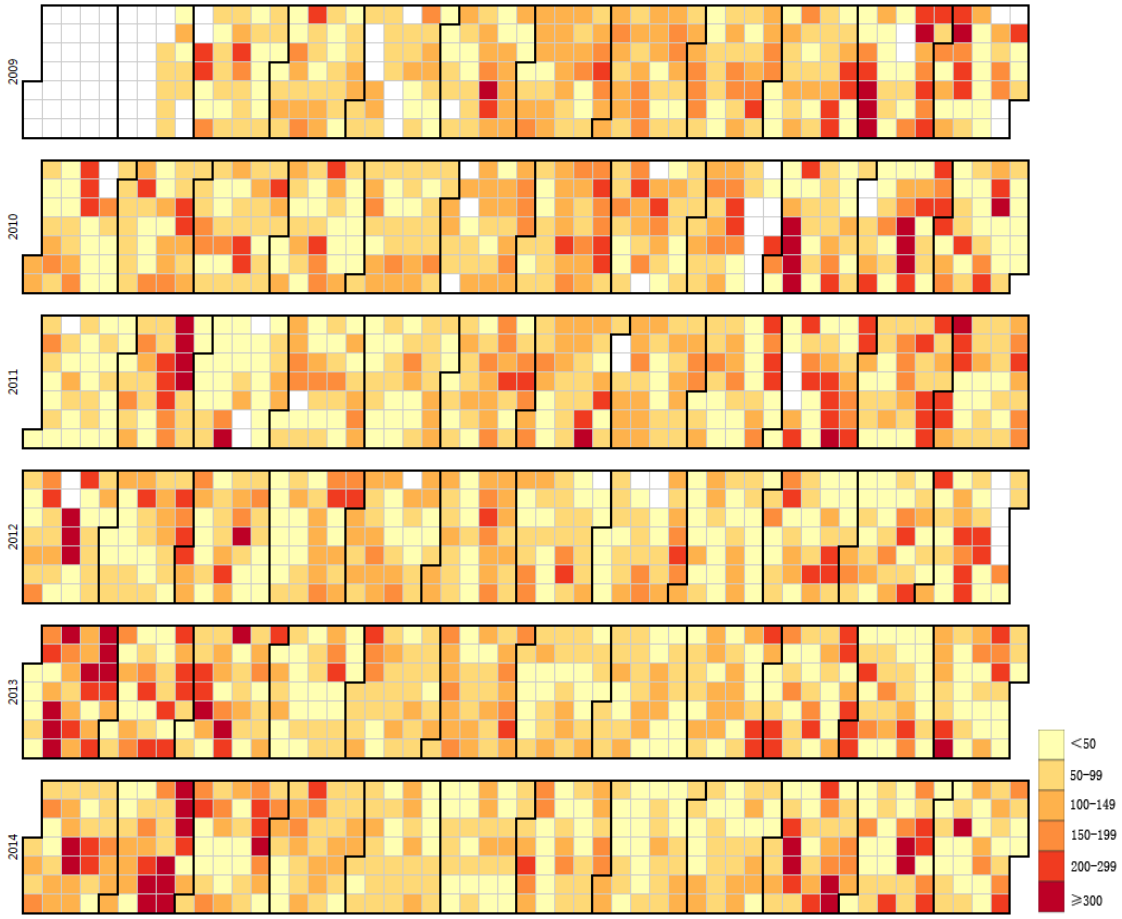


Figure 10. Calendar view: the relationship between pollutants and days from 2009 to 2014.

3.2.3. Spatial Characteristics

Spatial distribution of the mean AQI of $PM_{2.5}$ is analyzed in this section. Observations of the 36 stations from November 2013 to February 2014 are used. We first interpolate the AQI surface through the values of those stations by Ordinary Kriging (OK) [40], which is a geostatistical method where the weights for interpolation are computed by the neighboring values called "semivariances" (γ). In Equation (2), n is the number of pairs of sample points z separated by distance h, and $\hat{\gamma}(h)$ is the semivariogram which is a function of distance [41].

$$\hat{\gamma}(h) = \frac{1}{2n}\sum_{i=1}^{n}[z(x_i) - z(x_i + h)]^2 \tag{2}$$

The basic formula for OK is Equation (3), in which the λ_i is the kriging weight and $Z(x_0)$ is the observed value at point x_0.

$$\begin{cases} \hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \\ \lambda_i = var\left[\hat{Z}(x_0)\right] = E\left[\left\{\hat{Z}(x_0) - Z(x_0)\right\}^2\right] \end{cases} \tag{3}$$

With Equations (2) and (3), provided by Clark [42], we created the Geovisualization heatmap of Beijing air pollution shown in Figure 11. The interpolation heatmap presents higher concentrations of pollutants in red, and lower concentrations in blue. It can be observed that air pollution has an increasing trend from north to south. From the land use figure of the year 2011 [43], it can be found that the north area in Beijing is mainly wood and orchard land, the center area is urban land, and the

south is mainly arable land with a few residential areas. This could be further inspected through the satellite map of this area.

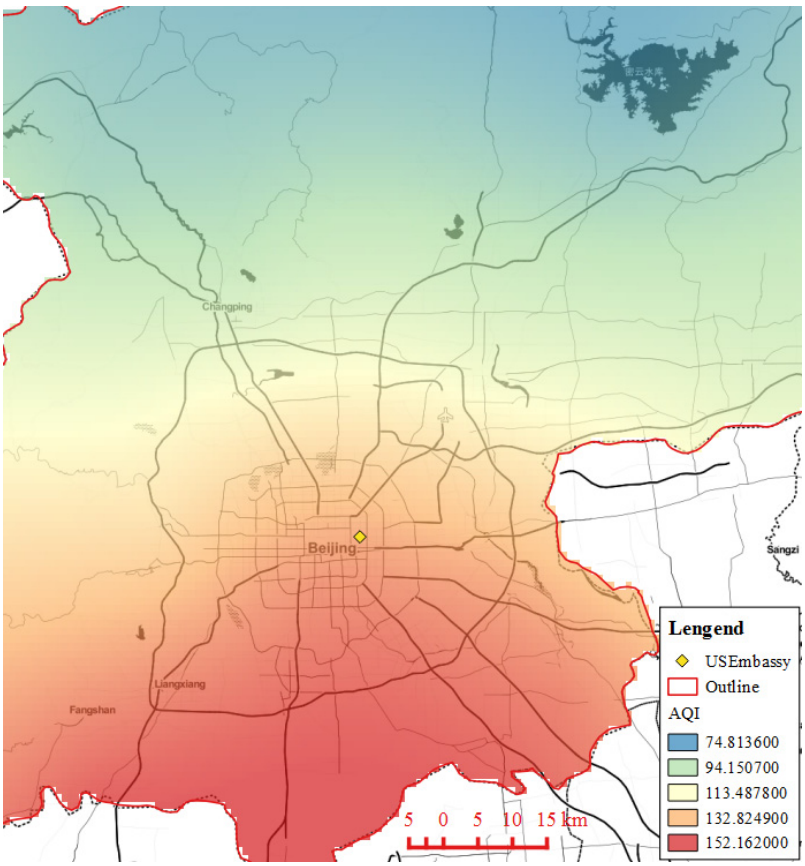


Figure 11. Average AQI Heatmap of $PM_{2.5}$ from November 2013 to February 2014 in Beijing.

A satellite map with 36 overlaid monitoring stations is shown in Figure 12. Stations 30, 31, and 32 are found in areas covered with vegetation. These stations showed lower concentrations of $PM_{2.5}$ compared with the other sites. The north and west of Beijing have higher terrain, are densely covered by vegetation, and are sparsely populated. Station 32 had the lowest pollutant concentration (Figure 6). As shown in Figure 12a, station 32 is surrounded by mountains and a lake and is far from residential areas. Figures 6 and 11 present that the central area (03–04, 06–16, and 20–22) has average air pollution concentrations. The southern areas of Beijing have a mix of arable and residential lands, and the air pollution concentrations in the corresponding stations (IDs 05, 17–19) have relatively higher AQIs. Stations 34–36 in the south end of Beijing, a major part of which is arable land, show the highest concentrations among other areas of the city. Figures in this section suggest that air pollutants exhibit strong spatial distribution. Vegetation, terrain, and land use influence air contamination. Vegetation can reduce the diffusion of atmospheric pollutants and absorb some pollutants. Arable lands have considerably higher air pollutant concentrations than vegetation-covered areas.

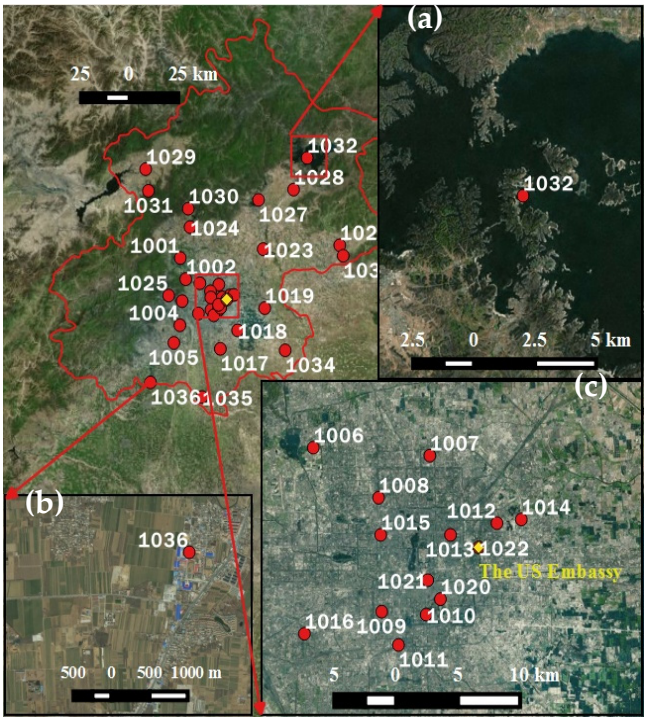


Figure 12. Distribution of monitoring stations (labels are IDs) overlaid on a satellite map. Three areas are enlarged to show the conditions of the corresponding land surface. (**a**) The lowest average observation was in station 32; (**b**) The highest one was in station 36; (**c**) The central area is also enlarged to show the locations of stations and the US Embassy, which is marked in yellow.

4. Conclusions

A visual exploration method was proposed to analyze air pollution data, which enables rapid processing and multi-perspective exploration of air pollution data to reveal spatio-temporal patterns and basic relationships among multiple variables. The developed method follows the observation–hypothesis–verification process of the scientific method and thus complies with the demands of empirical science. The proposed visual exploration method achieved rapid processing and accurate air pollution results for Beijing to guide the daily lives of residents and government decisions. Based on a series of multi-perspective visualizations of $PM_{2.5}$ data for Beijing, we conclude that the following propositions are suitable topics for further empirical study:

(1) A strong correlation existed between pollutants and wind speed. The positive correlation between pollutants indicates that $PM_{2.5}$, PM_{10}, and NO_2 may be emitted by the same sources, or one may be produced by the transformation of another through some type of chemical mechanism. Pollutants and wind speed were negatively correlated because wind accelerated the spread of pollutants and a mixture of fresh air may reduce pollutant concentration.

(2) Temporal characteristics were found through visual analytics. The average concentration of $PM_{2.5}$ in Beijing was greater than the AQI value of 50 over the six-year study period; therefore, residents suffer from long-term moderate pollution. The concentrations are more serious from October to March and from 6:00 p.m. to 4:00 a.m. on average. People should stay indoors during these time periods.

(3) Spatial distribution of air pollutants was also determined through geovisualization. Vegetation, terrain, and land use influenced air contamination. Vegetation could reduce the diffusion of atmospheric pollutants and absorb some pollutants. Arable lands had considerably higher concentrations of air pollutants than vegetation-covered areas.

The findings obtained in this study can be used as reference for further statistical analysis. By incorporating remote-sensing data, AQI data can be used to evaluate air quality at a regional or global extent. With the hot spots identified by our method, the severely polluted periods could be easily determined for air composition and transformation analysis from the physical and chemical perspectives; thus, the emission source could be more easily confirmed. Our future work will combine passive remote sensing and Mie/Raman LiDAR and incorporate extra models for more comprehensive analysis based on our previous studies [44]. Moreover, our method can be extended to visual analyses in other domains, such as soil pollution, climate change, and urban sprawl.

Acknowledgments: This work was supported by grants from the National Natural Science Foundation Projects of China (No. 41471323, PI: Hong Fan), the Science and Technology Development Project of Guizhou Province Tobacco Corporation of China National Tobacco Corporation (Contract No. 201407), and the National High Technology Research and Development Program of China (Grant No. 2012AA121401, PI: Jianya Gong). We sincerely thank Stephen C. McClure for the English modification of this paper.

Author Contributions: Huan Li conceived the idea, designed the experiments, and wrote the paper. Hong Fan provided important suggestions and supported the paper. Feiyue Mao partly analyzed the experimental data based on his knowledge of atmospheric science.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Open source visualization tools

The open source tools were used to realize the visualization in the paper. Because open source software is free, it is easily accessible by researchers in developing countries or young scientists who may not have enough funds to support their studies. In addition, the open codes are modifiable and extendable for custom usage. Moreover, open communities like Github, provide an abundance of example plots, which are updated in real-time for multi-perspective spatio-temporal visualization exploration. Therefore, it is reasonable and suitable to use those tools for visual analytics. In this appendix, we introduce two tools used for our paper: PivotTable.js and D3.js.

PivotTable.js

PivotTable.js [45] is an open source JavaScript library with drag-and-drop functionality based on jQuery and jQueryUI. PivotTable's core function is to import structured data as a summary table that displays specific attribute data with a drag-and-drop interface. The summary table can be rendered into various charts, transforming the table directly into figures. PivotTable has the following characteristics.

First, it is open source and lightweight. The core code is just a single file that compiles down to 6.3 kb of minified JavaScript that can run anywhere as long as jQuery and jQueryUI work.

Second, it can be used for formatted data extraction. Through PivotTable, the function $.pivotUtilities.derivers.dateFormat can reformat complicated date or date-time values into day, hour, minute, week, month values. Thus, time data can be easily extracted; for example, the time data format, "2/8/2013 9:10:30 PM", can be reformatted as shown in Table A1. This is a very useful function when working with temporal data. Derived attributes can be created on the fly based on the whole input record by passing it through a function.

Third, it is an efficient and intuitive basic visual analysis tool that includes the capacity to create heat maps and bar, line, and pie charts using interactive drag-and-drop operations.

Finally, users can easily use PivotTable to perform basic statistical analysis, such as calculating mean, sum and median. If necessary, customized TSV data can be outputted as data for inputting into the next layer for further visualization and analysis.

Figure A1 is the PivotTable GUI, showing its basic function areas. As seen in the figure, A and B are drop-down boxes that display options for users. A is the renderer, used to select the type of visualization, such as table, heat map, and line chart. B is the aggregator, used for selecting data for display, like mean AQI of $PM_{2.5}$. C is the container for the attributes, and D is the Row attribute. E is the Column, to which attributes are dragged from C for visualization. The attribute data can be

selected for analysis; for example, we can choose January from the 'month' attribute for analysis. F is the Data Visualization area, where the renderer builds an appropriate visualization.

PivotTable has data preprocessing capabilities. A global analysis visualization program can be used to preprocess data. By dragging and dropping attributes to the rows and columns and selecting TSV output in the renderer, the corresponding data are shown in the Data Visualization Area. This custom function makes it very easy and efficient to export data that can be directly used for further analysis applications.

Table A1. Formatting parameters and corresponding return values.

Parameter Interpolation	Return Value
%y: date.getFullYear ()	2013
%m: zeroPad (date.getMonth () +1)	2
%n: mthNames (date.getMonth ())	Feb
%d: zeroPad (date.getDate ())	8
%w: dayNames (date.getDay ())	Fri
%x: date.getDay ()	5 (Sunday is Zero)
%H: zeroPad (date.getHours ())	21 ("PM" is auto-computed)
%M: zeroPad (date.getMinutes ())	10
%S: zeroPad (date.getSeconds ())	30

Figure A1. PivotTable Graphical User Interface (GUI). A to F represent function areas.

D3.js

Data-Driven Documents (D3) [46] is a JavaScript library that uses HTML, SVG and CSS techniques to visualize data. Similar to PivotTable, it is a lightweight library with a core functionality file size less than 150 kb. It is data-driven for data inspections and operations based on the Document Object Model (DOM). D3 has detailed API documentation and an abundance of examples, and its maintenance and updating frequency is high. Thus, D3 is very suitable for statistical visualization.

Appendix B. Data quality check of the U-Air data

It has been noted in the paper that the U-Air data used in the cased study has some defects. This supplement file is to strengthen the idea and give more inspection to the original data. We use another visualization to show the data missing and a table to present the anomaly weather data in it.

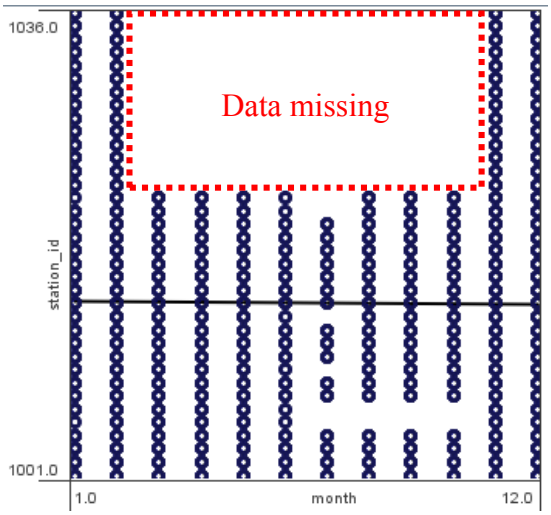


Figure A2. A scatterplot for 12 months and 36 stations.

In Figure A2, the red dot rectangle marks the missing data area. The black line is the regression line of the observations. The scatter plot is another kind of visual methods for missing data check. It confirms the results shown in Figure 3 in Section 3.1, identifying that data are missing from some stations from March to October.

As seen in Figure A3, snowy weather occurred from April to June, showing incorrect records because this is impossible in Beijing (39.92N, 116.44E); there was no sunny weather from June to October, also suggesting data anomalies. Considering the defects identified in this figure, we decided to replace the weather records in the U-Air data.

weather name / month name	snowy	cloudy	sunny	overcast	rainy	foggy	dusty	Totals
Jan	168.05	159.78	115.11	254.66		336.05		**131.36**
Feb	208.22	235.61	97.55	180.08	251.96	185.37		**142.32**
Mar	188.41	268.25	198.44	132.49	258.91		150.41	**217.48**
Apr	77.03	125.02	89.92	136.13	175.86			**118.33**
May	114.43	181.92	148.98	154.04	147.53			**149.03**
Jun	114.14	163.34		174.94	212.09			**182.84**
Jul		123.33		144.31	133.69			**129.80**
Aug		101.71		161.26	147.21			**125.13**
Sep		140.27		168.41	155.33	229.17		**145.76**
Oct		148.27		208.18	180.09	301.29		**157.06**
Nov		210.95	79.65	187.85	228.28	178.49		**89.48**
Dec		83.27	98.60	222.48		254.69		**104.72**
Totals	**149.62**	**141.22**	**103.23**	**171.41**	**158.59**	**249.13**	**150.41**	**134.12**

Figure A3. The average AQI of $PM_{2.5}$ as shown in PivotTable; rows represents months, and columns contain the weather data. With PivotTable.js, this table can be reproduced by making the row as "month name", and the column with "weather name".

Appendix C. Data processing

This part focuses on the design for visualization and data preprocessing. Through the, we developed three hypotheses and a corresponding visualized design and implementation tools to test these propositions. Data preprocessing is also considered.

Table A2. Hypothesis and visualization project design for verification.

Hypotheses	Plots type	Visualization tools	Data Preprocessing
Relationship exists among pollutants and wind speed	Scatter plots	D3.js	By PivotTable.js
There are some regular patterns in time	Heat maps (Circular heat chart, Calendar view)	D3.js	
Concentration distributionof pollutants in space	Geovisualization	Openlayers	

Three hypotheses are shown in Table A2. The first hypothesis is that there is a correlation among the AQI of $PM_{2.5}$, PM_{10}, NO_2 and wind speed. This can be visualized by scatterplots showing the relationship between any two factors. The second is that there is a regular time pattern for air pollutants. Heat maps would be an ideal visualization method, including a circle heat map, calendar view, and heat matrix, as generated using D3. The third hypothesis is that air contaminants have a geographical distribution. This can be represented appropriately by geovisualization, a method suitable for illustrating continuous spatial distributions. There are different ways to achieve Geovisualization with D3 or OpenLayers.

Data preprocessing precedes visualization. Data pretreatment in PivotTable is efficient because of its built-in deriver generators. After preprocessing, data can be used for visualization. As analyzed in Appendix B, weather data has some anomalies, so we replace that part with rainy weather in which the precipitation in mm [36]. With this correction, we can analyze the correlation between the severity of air pollution and the precipitation.

The data production for Figures 9 and 10 in the paper is by PivotTable with USE data. The Figure 9a is produced by setting rows with "Year" and "Month", while Figure 9b is produced by setting rows with "Year" and "Hour", both choosing "TSV Export" in the Aggregator of PivotTable. Then the data in Tab separated values (TSV) form can be easily visualized by D3.js without additional data processing.

References

1. Mao, F.; Duan, M.; Min, Q.; Gong, W.; Pan, Z.; Liu, G. Investigating the impact of haze on modis cloud detection. *J. Geophys. Res. Atmos.* **2015**, *120*, 12237–12247. [CrossRef]
2. Zhang, Y.L.; Cao, F. Fine particulate matter ($PM_{2.5}$) in China at a city level. *Sci. Rep.* **2015**, *5*, 14884. [CrossRef] [PubMed]
3. Matejicek, L.; Engst, P.; Jaňour, Z. A GIS-based approach to spatio-temporal analysis of environmental pollution in urban areas: A case study of Prague's environment extended by LiDAR data. *Ecol. Model.* **2006**, *199*, 261–277. [CrossRef]
4. Retalis, A.; Sifakis, N. Urban aerosol mapping over Athens using the differential textural analysis (dta) algorithm on meris-envisat data. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 17–25. [CrossRef]
5. Chen, W.; Tang, H.; Zhao, H. Diurnal, weekly and monthly spatial variations of air pollutants and air quality of beijing. *Atmos. Environ.* **2015**, *119*, 21–34. [CrossRef]
6. Carslaw, D.C.; Ropkins, K. Openair—An R package for air quality data analysis. *Environ. Model. Softw.* **2012**, *27–28*, 52–61. [CrossRef]
7. Janssen, N.A.H.; Fischer, P.; Marra, M.; Ameling, C.; Cassee, F.R. Short-term effects of $PM_{2.5}$, PM_{10} and $PM_{2.5-10}$ on daily mortality in the Netherlands. *Sci. Total Environ.* **2013**, *463*, 20–26. [CrossRef] [PubMed]

8. Shi, W.; Wong, M.S.; Wang, J.; Zhao, Y. Analysis of airborne particulate matter ($PM_{2.5}$) over Hong Kong using remote sensing and GIS. *Sensors* **2012**, *12*, 6825–6836. [CrossRef] [PubMed]
9. Van Wijk, J.J. The value of visualization. *Proc. IEEE* **2005**. [CrossRef]
10. Cai, Z.; Wang, Q.; Weng, M.; Jiang, S.; Du, Q. Information organization and visualization mechanism of electronic map. *Geo-Spat. Inf. Sci.* **2008**, *11*, 262–268. [CrossRef]
11. Andrienko, G.; Andrienko, N.; Demsar, U.; Dransch, D.; Dykes, J.; Fabrikant, S.I.; Jern, M.; Kraak, M.J.; Schumann, H.; Tominski, C. Space, time and visual analytics. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 1577–1600. [CrossRef]
12. Kraak, M.J. Geovisualization illustrated. *ISPRS J. Photogramm. Remote Sens.* **2003**, *57*, 390–399. [CrossRef]
13. Keim, D.; Andrienko, G.; Fekete, J.-D.; Görg, C.; Kohlhammer, J.; Melançon, G. Visual analytics: Definition, process, and challenges. In *Information Visualization*; Springer: Berlin/Heidelberg, Germany, 2008.
14. Li, X.; Coltekin, A.; Kraak, M.-J. Visual exploration of eye movement data using the space-time-cube. In *Geographic Information Science*; Fabrikant, S.I., Reichenbacher, T., van Kreveld, M., Schlieder, C., Eds.; Springer: Heidelberg, Germany, 2010; Volume 6292, pp. 295–309.
15. Zirui, L.; Bo, H.; Dongsheng, J.; Yonghong, W.; Mingxing, W.; Yuesi, W. Diurnal and seasonal variation of the $PM_{2.5}$ apparent particle density in Beijing, China. *Atmos. Environ.* **2015**, *120*, 328–338.
16. Pope, C.A.; Burnett, R.T.; Thun, M.J.; Calle, E.E.; Krewski, D.; Ito, K.; Thurston, G.D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama-J. Am. Med. Assoc.* **2002**, *287*, 1132–1141. [CrossRef]
17. Revised Air Quality Standards for Particle Pollution and Updates to the Air Quality Index (AQI). Available online: http://www.epa.gov/airquality/particlepollution/2012/decfsstandards.pdf (accessed on 25 February 2016).
18. Technical Regulation on Ambient Air Quality Index (on Trial). Available online: http://kjs.mep.gov.cn/hjbhbz/bzwb/dqhjbh/jcgfffbz/201203/W020120410332725219541.pdf (accessed on 25 February 2016).
19. Jeon, S.J.; Meuzelaar, H.L.C.; Sheya, S.A.N.; Lighty, J.S.; Jarman, W.M.; Kasteler, C.; Sarofim, A.F.; Simoneit, B.R.T. Exploratory studies of PM_{10} receptor and source profiling by GC/MS and principal component analysis of temporally and spatially resolved ambient samples. *J. Air Waste Manag. Assoc.* **2001**, *51*, 766–784. [CrossRef] [PubMed]
20. Chu, D.A.; Kaufman, Y.J.; Zibordi, G.; Chern, J.D.; Mao, J.; Li, C.C.; Holben, B.N. Global monitoring of air pollution over land from the earth observing system-terra moderate resolution imaging spectroradiometer (MODIS). *J. Geophys. Res.-Atmos.* **2003**, *108*. [CrossRef]
21. Wang, J.; Christopher, S.A. Intercomparison between satellite-derived aerosol optical thickness and $PM_{2.5}$ mass: Implications for air quality studies. *Geophys. Res. Lett.* **2003**, *30*. [CrossRef]
22. Othman, N.; MatJafri, M.Z.; Lim, H.S.; Abdullah, K. Retrieval of aerosol optical thickness (AOT) and its relationship to air pollution particulate matter (PM_{10}). In Proceedings of the Sixth International Conference on Computer Graphics, Imaging and Visualization, 2009 CGIV '09, Tianjin, China, 11–14 August 2009; pp. 516–519.
23. Zheng, Y.; Liu, F.; Hsieh, H.-P. U-air: When urban air quality inference meets big data. In *19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: Chicago, IL, USA,, 2013; pp. 1436–1444.
24. Bostock, M.; Ogievetsky, V.; Heer, J. D-3: Data-driven documents. *IEEE Trans. Vis. Comput. Graph.* **2011**, *17*, 2301–2309. [CrossRef] [PubMed]
25. Matejicek, L. Spatio-temporal analysis of environmental pollution in urban areas: A case study of the environment in the city of Prague. In Proceedings of the 19th International Congress on Modelling and Simulation (Modsim2011), Perth, Australia, 12–16 December 2011; pp. 1909–1915.
26. Guo, D.S.; Chen, J.; MacEachren, A.M.; Liao, K. A visualization system for space-time and multivariate patterns (Vis-Stamp). *IEEE Trans. Vis. Comput. Graphics* **2006**, *12*, 1461–1474.
27. Andrienko, G.; Andrienko, N. A general framework for using aggregation in visual exploration of movement data. *Cartogr. J.* **2010**, *47*, 22–40. [CrossRef]
28. Kveladze, I.; Kraak, M.J.; van Elzakker, C.P.J.M. A methodological framework for researching the usability of the space-time cube. *Cartogr. J.* **2013**, *50*, 201–210. [CrossRef]
29. Li, H.; Fan, H.; Wu, H.; Feng, H.; Li, P. Resdap: A real-time data provision system architecture for sensor webs. In *Web and Wireless Geographical Information Systems*; Springer: Heidelberg, Germany; pp. 85–99.

30. Fan, H.; Li, H. An on-Demand provision model for geospatial multisource information with active self-adaption services. *Proc. SPIE* **2015**, *9815*. [CrossRef]
31. Fan, M.; Fan, H.; Chen, N.; Chen, Z.; Du, W. Active on-demand service method based on event-driven architecture for geospatial data retrieval. *Comput. Geosci.* **2013**, *56*, 1–11. [CrossRef]
32. Li, L.; Losser, T.; Yorke, C.; Piltner, R. Fast inverse distance weighting-based spatiotemporal interpolation: A web-based application of interpolating daily fine particulate matter $PM_{2.5}$ in the contiguous U.S. Using parallel programming and k-d tree. *Int. J. Environ. Res. Public Health* **2014**, *11*, 9101–9141. [CrossRef] [PubMed]
33. Van Wijk, J.J.; Van Selow, E.R. Cluster and calendar based visualization of time series data. *IEEE Symp. Inf. Vis.* **1999**, *140*. [CrossRef]
34. Department of State Air Quality Monitoring Program. Available online: http://www.stateair.net/web/mission/1/ (accessed on 10 December 2015).
35. Zheng, Y. Urban Air. Microsoft. Available online: http://research.microsoft.com/en-us/projects/urbanair/default.aspx (accessed on 15 October 2015).
36. World Weather. Available online: http://en.tutiempo.net/ (accessed on 20 January 2016).
37. Qiao, Z.; Tian, G.; Xiao, L. Diurnal and seasonal impacts of urbanization on the urban thermal environment: A case study of Beijing using MODIS data. *ISPRS J. Photogramm. Remote Sens.* **2013**, *85*, 93–101. [CrossRef]
38. Gong, W.; Zhang, T.; Zhu, Z.; Ma, Y.; Ma, X.; Wang, W. Characteristics of $PM_{1.0}$, $PM_{2.5}$, and PM_{10}, and their relation to black carbon in Wuhan, central China. *Atmosphere* **2015**, *6*, 1377–1387. [CrossRef]
39. Logan, T.; Xi, B.; Dong, X. A comparison of the mineral dust absorptive properties between two Asian dust events. *Atmosphere* **2013**, *4*, 1–16. [CrossRef]
40. Cressie, N. The origins of kriging. *Math. Geol.* **1990**, *22*, 239–252. [CrossRef]
41. Sanabria, L.; Qin, X.; Li, J.; Cechet, R.; Lucas, C. Spatial interpolation of McArthur's forest fire danger index across Australia: Observational study. *Environ. Model. Softw.* **2013**, *50*, 37–50. [CrossRef]
42. Clark, I. *Practical Geostatistics*; Applied Science Publishers: London, UK, 1979; Volume 3.
43. Huang, Y.; Nian, P.; Zhang, W. The prediction of interregional land use differences in Beijing: A markov model. *Environ. Earth Sci.* **2015**, *73*, 4077–4090. [CrossRef]
44. Wang, W.; Gong, W.; Mao, F.; Zhang, J. Long-term measurement for low-tropospheric water vapor and aerosol by Raman LiDAR in Wuhan. *Atmosphere* **2015**, *6*, 521–533. [CrossRef]
45. PivotTable.js. Available online: https://github.com/nicolaskruchten/pivottable (accessed on 24 October 2015).
46. D3.js. Available online: http://d3js.org/ (accessed on 14 November 2015).