CrossMark

REGULAR PAPER

Wenting Zhang · Yinqiao Wang · Qiong Zeng · Yunhai Wang · Guoning Chen · Tao Niu · Changhe Tu · Yi Chen

Visual analysis of haze evolution and correlation in Beijing

Abstract Haze is a hazardous atmospheric phenomenon that threatens human health and leads to severe economic problems. A number of weather factors are relevant to the emergence and evolvement of haze. In this paper, we present a visual analytics system for haze study, including its evolution and correlations to a number of weather factors. Specifically, we introduce a haze event detection algorithm based on common haze identification rules in meteorology using the $PM_{2.5}$ concentration data. We develop a comparative visualization to consistently overview trends of scalar variables and wind directions, in which wind patterns are extracted via clustering streamlines at user-given sampling time. To study the correlation between wind and $PM_{2.5}$, we decompose time steps into time intervals according to the temporal similarity of streamlines. Additionally, we develop a 1D function dissimilarity measurement to study the temporal correlation between $PM_{2.5}$ concentrations and relevant weather factors, such as wind strength, relatively humidity and planetary boundary layer. Furthermore, we employ particle advection using pathline computation within the wind field to locate the origins and destinations of particles seeded in user-interested areas. We applied our system to study of a number of hazes occurring in January 2013 in Beijing, (China). Interpretations and evaluations from domain experts demonstrate the effectiveness of our system in facilitating haze study.

Keywords Haze study · Visual analysis · Streamline

1 Introduction

Haze is an atmospheric phenomenon in which small dust, smoke as well as salt and sand particles are mixed with water vapor. This reduces visibility to below 10 km (Houze 2014) (see Fig. 1). Haze typically occurs near the ground, i.e., at the bottom of the atmosphere. Two main negative effects are reported: First, haze is mainly comprised of dust and other aerosol particles that contains sulfuric acid, nitric acid, and

Wenting Zhang and Yinqiao Wang have contributed equally to this work and joint first authorship.

W. Zhang · Y. Wang · Q. Zeng (✉) · Y. Wang · C. Tu
Shandong University, Qingdao, China
E-mail: qiong.zn@gmail.com

G. Chen
University of Houston, Houston, USA

T. Niu
Chinese Academy of Meteorological Sciences, Beijing, China

Y. Chen (✉)
Beijing Technology and Business University, Beijing, China
E-mail: chenyi@th.btbu.edu.cn

Fig. 1 Haze is an atmospheric phenomenon, which reduces visibility and influences human activities. Here shows an example of severe haze in Beijing on January 11, 2013 (http://pharm.vogel.com.cn/2013/0306/news_350750.html)

hydrocarbons (Wang et al. 2006). When haze occurs, these aerosol particles are aggregating close to the ground, while significantly increasing air pollution to a level that is harmful to human health. Secondly, the obscuration resulted by haze heavily influences people's activities and thus might harm national economy. As an example, on February 17, 1994, a severe haze struck Beijing and reduced visibility to below 50 m, which enforced closing the international airport for more than 30 h. More than 250 flights were canceled or delayed, and more than 16,000 passengers were stranded.

Particles contained in the haze have varying sizes. Among them, $PM_{2.5}$ particles with diameters less than or equal to 2.5 microns are the most noteworthy group. Such particles can be suspended in the air for a long time. Therefore, $PM_{2.5}$ is a typical factor to represent haze in meteorology. In general, a higher $PM_{2.5}$ concentration indicates a more severe air pollution level in the haze. Additionally, $PM_{2.5}$ particles are extremely chemically active, and thusmore easily adhere to toxic and hazardous substances (Sun et al. 2006). Therefore, $PM_{2.5}$ particles and their concentration in the air have direct impact on human health and environmental quality. Understanding the evolution of $PM_{2.5}$ and its correlations with other weather conditions (such as temperature, humidity and wind) will provide valuable insights into the analysis of haze event and prediction of future haze events. This will enable local governments and residents to plan accordingly and to potentially reduce negative effects. However, there is no existing visual system for haze analytics.

In this work, we develop a visual analytics system that enables experts to identify haze events and to study how its evolution is influenced by different weather conditions. In our visual system, we leverage the evolution of $PM_{2.5}$ particles to trace haze event, as the same as haze analytics in meteorology. More specifically, we first develop an effective technique to identify the key times when a haze event starts (upward, stabilizing, downward stages) based on the variation of the $PM_{2.5}$ concentration. Our method can be used to detect multiple $PM_{2.5}$ events in a given time period. Next, we inspect other weather conditions around the identified key times of $PM_{2.5}$ event and try to understand how they contribute to the event by analyzing a number of temporal and spatial correlations. Rather than simply calculating the correlation coefficients between variables over time, we explore how these variables contribute to different stages of haze. Specifically, we analyze how wind and relative humidity influence the evolution of $PM_{2.5}$. In order to provide a visual summary of these variables, we design a comparative visualization system that allows to show trends of multiple scalar and vector variables (such as wind) together. To help domain experts to analyze where a high $PM_{2.5}$ concentration comes from, we also analyze the spatial correlation between haze in Beijing and its nearest provinces by using pathline advection.

The main contributions in this work include:

- A novel visual system for haze event detection and its correlations to various weather variables, by integrating scalar and vector field analysis.
- A haze detection algorithm, a phase correlation computation scheme and a vector correlation method have been proposed.
- Case studies to demonstrate its usefulness using haze data in Beijing City based on the feedback from domain experts.

Data In January 2013, eastern China experienced a strong and long-lasting haze event. The number of consecutive days with haze in Beijing and nearby areas broke the record since 1961. As the capital of China, Beijing and its air quality have received substantial attention around the world. Therefore, this study will concentrate on haze events occurring in Beijing and its neighboring areas. In particular, we select haze simulation data generated by the WRF model (Michalakes et al. 2004), for January 2013 and October 2014, which includes 23 vertical layers, each layer having 29 variables, the ground layer having additional 9 variables. The $PM_{2.5}$ concentration exhibits large variations during these two periods. The data is sampled every hour during these two periods where the size of each sample is around 0.5 GB. Under the guidance of our domain collaborators, we focus on five variables that they are mostly interested in. The five variables are the fundamental meteorological attributes, including $PM_{2.5}$ concentration, wind, relative humidity, temperature and planetary boundary layer in our study. The first four variables come from the second layer, and the last one comes from the ground layer.

The rest of the paper is organized as follows: After briefly introducing the related work in Sect. 2, we give an overview of our system in Sect. 3. We then describe haze event detection algorithm in Sect. 4 and discuss the temporal correlation between $PM_{2.5}$ concentration with scalar variables and wind in Sects. 5 and 6, respectively. Finally, we present case studies with domain experts and finish with conclusion and future directions.

2 Related work

Weather data visualization In the past decades, the visualization of meteorological (and climate) data has become an important topic (Hibbard 1986; Treinish 1994; Wang et al. 2015). Many practical visualization tools have been developed, such as Vis5D (Hibbard and Paul 1998; Hankin et al. 1996; Berman et al. 2001). However, there is still a gap between the capabilities of advanced visualization systems and the domain work in climate research (Nocke et al. 2008; Tominski et al. 2011). To support hypothesis generation from large-scale climate data, Kehrer et al. (2008) and Ladstadter et al. (2010) propose a novel visualization pipeline. Janicke et al. (2009) take advantage of wavelet analysis in visualizing climate variability changes. Kehrer et al. (2011) propose an interface for heterogeneous scientific data analysis, which is also applicable for meteorological data analysis. Since simulated weather data often are generated by ensemble simulation, many ensemble data visualization and analysis techniques (Guo et al. 2013; Potter et al. 2009; Sanyal et al. 2010) have been applied to evaluate climate models or study their uncertainty. All these works support domain experts in understanding the atmospheric state to a certain extent and contribute to many practical applications.

There exist several visual analytic systems for air pollution problems (Li et al. 2016; Qu et al. 2007; Zhou et al. 2017). Qu et al. (2007) present an integrated system to visualize air pollution attributes with polar system, parallel coordinates and the weighted complete graph. They mainly focus on attribute correlations, regional comparisons and pattern estimation. Li et al. (2016) propose a visualization system with four interrelated views to analyze smogs at different spatial and sequential scales, aiming to help domain experts qualitatively discover correlations of meteorological attributes and intrinsic patterns. However, to our best knowledge, a comprehensive visual analytics system that supports understanding of causes and evolution of hazes with the combination of scalar and vector fields does not yet exist.

Vector field visualization Within climate studies, a wind system is typically modeled as an *unsteady* (or *time-dependent*) vector field $V(\mathbf{x}, t)$ defined in a spatio-temporal domain as a function of time. The trajectory of a *massless* particle starting at $\mathbf{x}$ and at time t_0 under $V(\mathbf{x}, t)$, i.e., its *pathline*, is computed by:

$$\mathbf{p}_{\mathbf{x},t_0}(t) = \mathbf{x} + \int_0^t V(\mathbf{p}_{\mathbf{x},t_0}(\tau), t_0 + \tau) \mathrm{d}\tau. \tag{1}$$

The instantaneous vector field at any given time is a steady vector field, $V(\mathbf{x})$. The patterns of this vector field are typically revealed by its streamlines that can be defined and computed as follows:

$$\mathbf{p}_{\mathbf{x}}(t) = \mathbf{x} + \int_0^t V(\mathbf{p}_{\mathbf{x}}(\tau)) \mathrm{d}\tau. \tag{2}$$

Streamline placement is a popular technique for visualizing steady vector fields (McLoughlin et al. 2010). To reduce cluttering of densely placed streamlines and to reveal salient patterns of vector fields, streamline clustering techniques are used (Ferreira et al. 2013; Lu et al. 2013; Wang et al. 2017; Yu et al. 2012). They are also employed to classify streamlines into different clusters based on certain distance metrics (McLoughlin et al. 2013). Pathlines and their computation have also been widely used for the

interpretation of various unsteady vector fields (Guo et al. 2014; Pobitzer et al. 2012; Shi et al. 2009). In this work, we adapt and modify the previously proposed spatial streamline clustering technique by Yu et al. (2012) for studying the correlation between wind fields and haze events, and introduce a temporal streamline clustering technique. We also utilize pathline computation to study the advection of haze particles before, during and after a haze event.

Haze study Investigation of sources, formation and transport of haze is an important topic in meteorologic research (Malm 1992; Schichtel et al. 2001). By studying patterns and trends of haze over the USA for the period of 1980–1995, Schichtel et al. (2001) found that the reduction in haze was consistent with the reduction in $PM_{2.5}$. Chen et al. (2003) investigate summertime haze formation in the mid-Atlantic region by analyzing changes in the components in $PM_{2.5}$. Kang et al. (2004) evaluate the chemical characteristics of acidic gas pollutants and $PM_{2.5}$ during hazy episodes and investigate where high $PM_{2.5}$ concentrations come from. With the rapid urbanization and motorization, haze episodes frequently occurred in Beijing in recent years. Among them, the severe haze episode occurred in January 2013 last around a whole month. Many studies (Wang et al. 2014b; Sun et al. 2014; RenHe et al. 2014) have been carried out to investigate the pattern of this haze episode. Almost all of these methods analyze the formation mechanisms by using automatic statistical methods to characterize the related conditions of air pollution. In this paper, we propose a visual analysis method to analyze the correlation between $PM_{2.5}$ concentrations and other related meteorologic variables such as wind, relative humidity and planetary boundary layer. Moreover, our pathline advection enables meteorologists to interactively explore the origin and development of $PM_{2.5}$ particles involved in haze events.

3 System overview

Studying haze events consists of two steps: (1) detection of haze events and (2) correlation analysis of $PM_{2.5}$ concentration with other weather factors.

Since haze can be characterized by $PM_{2.5}$ concentration (Schichtel et al. 2001), our haze identification is achieved by analyzing the spatial distribution of $PM_{2.5}$ concentration over time. For haze event detection, we aim at identifying a complete event including its starting time, breaking out and ending time. We first represent the $PM_{2.5}$ concentration in Beijing as 1D curve (with respect to time, cf. Fig. 3). We then devise an effective algorithm based on the knowledge of domain experts and a corresponding 1D function analysis to identify individual haze events. (Sect. 4).

Next, we study the correlation between $PM_{2.5}$ concentration and a number of other weather factors that are of interest for domain experts. They are particularly interested in how planetary boundary layer (PBL) (Haby 2010), wind fields and relative humidity (RH) influence the evolution of hazes. These weather factors can be classified into two types, i.e., *scalar quantities*, such as PBL, wind strength, and RH as well as *vector-valued quantities*, such as wind direction. For processing the scalar quantities, we conduct a phase correlation analysis to estimate the delay between different weather processes. We then apply both the conventional correlation analysis using the Pearson Product-Moment Correlation Coefficient (PPMCC) (Stigler 1989) and the proposed phase correlation to estimate the influence of different weather factors to the process of haze formation during a haze event (Sect. 4).

For the vector-valued quantities, we perform a spatial clustering of streamlines to summarize the overall wind pattern at any given time. This enables us to create a comparative visual analytics tool with the abovementioned scalar quantities (Sect. 6.1). We then study the temporal influence of a wind field to the $PM_{2.5}$ concentration via a temporal streamline clustering process (Sect. 6.2). Finally, we perform a spatial correlation using particle advection based on the wind fields to understand the origins and destinations of $PM_{2.5}$ particles within a specific haze event (Sect. 6.3).

Figure 2 provides a snapshot of our visual analytics system. It consists of three major views (1–3) and a control panel (4). View (1) shows the comparative visualization of a variety of scalar variables over time (bottom) as well as the summarized wind patterns (top). View (2) shows a wind field visualization at a user-specified time using streamlines. The background color shows the spatial distribution of the $PM_{2.5}$ concentration. In particular, we normalize this distribution to the range [0, 1], and visualize it on top of the geographical map of the area using alpha color blending. We use a white to yellowish-gray (RGB = (220,189,97)) color coding with zero mapped to white and the largest $PM_{2.5}$ value mapped to yellowish-gray. View (3) provides a color plot of the selected scalar quantity at the same user-specified time. Here, the relative humidity is shown with a rainbow color coding (blue indicating low and red for high). The three major views integrate our analysis on the five variables and show them in a single interface. Domain experts can explore the system further by tuning parameters in the control panel.
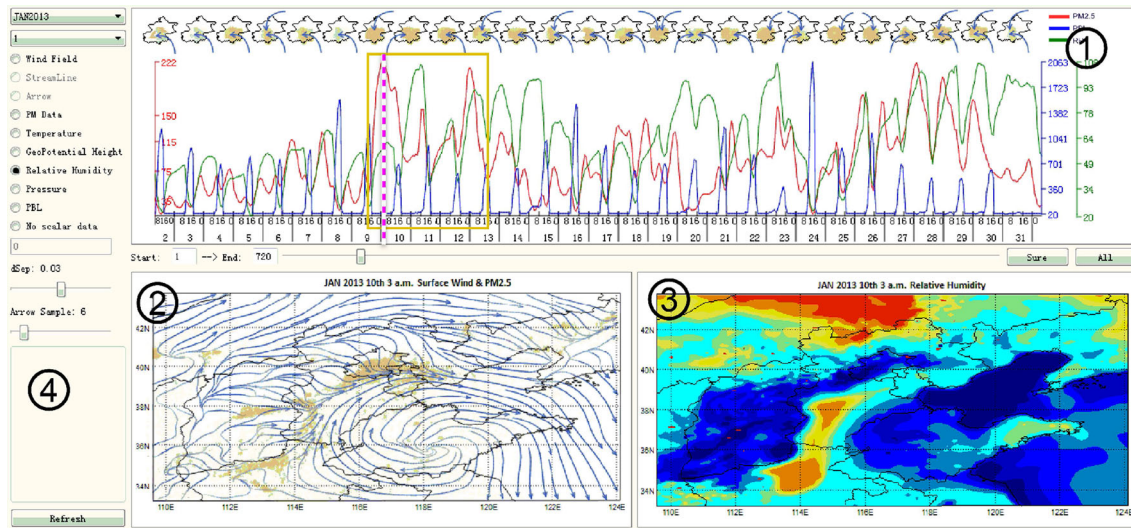
Fig. 2 Interface of our system, which consists of three major views (1–3) and a control panel (4). The time range covered by a yellow box in view (1) indicates a haze event and the pink dash line selects a time point where the corresponding variables of haze concentration and wind are shown in view (2), relative humidity is shown in view (3)

4 Haze event detection

In a preprocessing step, we create the PM$_{2.5}$ concentration curve from the simulated values obtained from the discrete spatial sampling of the Beijing area. Figure 3a shows the corresponding plot for January 2013 in Beijing. Each value is obtained by averaging all PM$_{2.5}$ concentration values for the sampling positions in
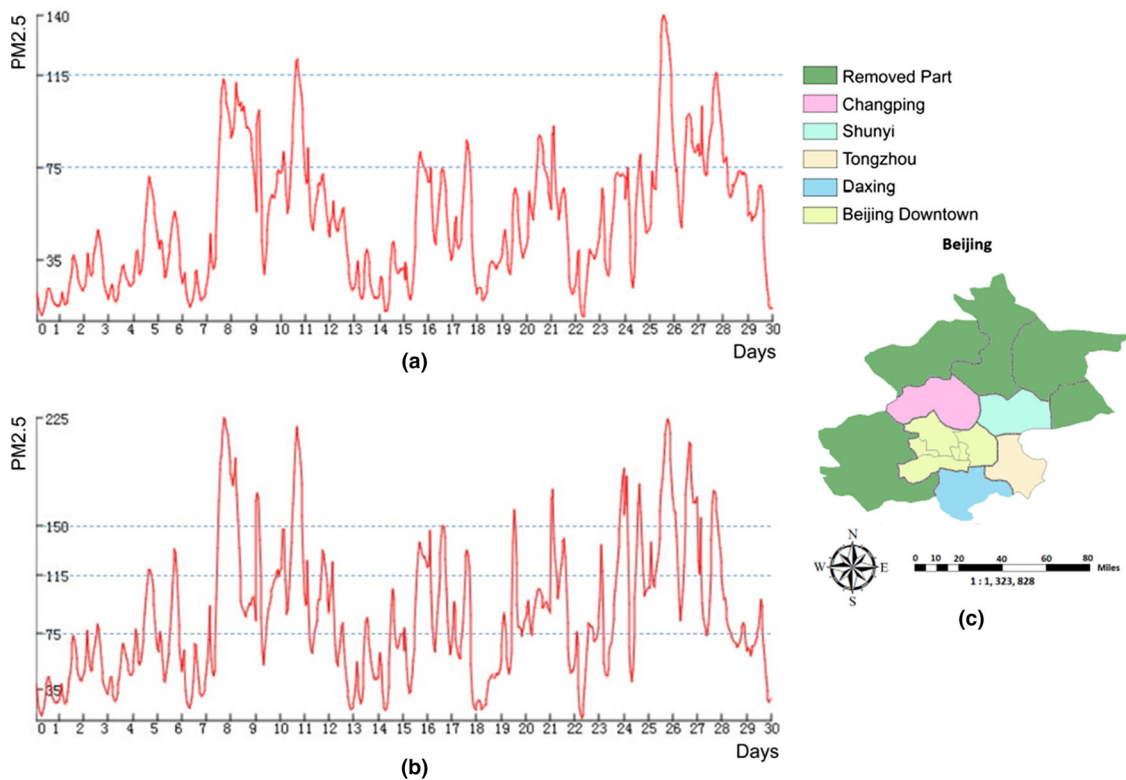

Fig. 3 PM$_{2.5}$ summarization for Beijing with all sampling points (**a**) and samples falling solely in the urban area (**b**), where the non-urban area is shown in green in **c**

Beijing area. Figure 3b provides a similar PM$_{2.5}$ plot, but is computed by removing the samples that are from the mountain (i.e., non-urban) regions around Beijing. Such locations are indicated by the green areas in Fig. 3c. As can be seen, after removing non-urban regions from the PM$_{2.5}$ summarization, the actual PM$_{2.5}$ level in Beijing City is revealed, which is much more severe than when looking at the entire area.

4.1 Haze identification criteria

After this preprocessing step, we perform a haze event detection based on the PM$_{2.5}$ concentration plot. Looking at the air quality forecast of China (government 2012), one important threshold that is used to determine "Good" air quality or "Slightly Polluted" is a value of 75 for the PM$_{2.5}$ concentration. If the PM$_{2.5}$ value is below 75, the air quality is generally considered to be good in China, otherwise, the air is polluted. We will use this threshold to locate certain key points within the PM$_{2.5}$ plot that may indicate the beginning or ending of a haze period. Another criterion for identification haze using PM$_{2.5}$ concentration is that a complete haze event should cover a time period for at least 24 h, and within this time period the PM$_{2.5}$ concentration should be larger than 75; furthermore, the peak value should be at least 115.

However, during a haze event, the PM$_{2.5}$ concentration may exhibit some large fluctuations and its value could temporarily be smaller than 75. A detailed explanation for this fluctuation will be provided in Sect. 7.2. If the dropping for this value lasts for less than 24 h, it is still considered to be part of the same haze event. Figure 4 shows an example for detected haze events based on the PM$_{2.5}$ concentration plot. Five haze events were detected based on the above criteria (highlighted by the boxes in yellow). Within each detected haze, the fluctuation of the PM$_{2.5}$ value is obvious. For many times during these periods, the value has dropped below the 75 threshold. However, since each of these droppings occurred during a very short period, e.g., $dt < 12$ h, the corresponding curves are still part of the haze events. Similarly, the ending time of a haze is the time when the value finally drops below 75. On the other hand, even though a number of PM$_{2.5}$ concentrations during days 14–17 were larger than 75, they are not deemed to be part of a haze event since none of these peaks reached 115.

4.2 Detection algorithm

Based on the above criteria, we now describe our algorithm for haze event detection within the PM$_{2.5}$ plot, (called $f(q_i)$) where the q_i are the sampled data points.

Step 1 We perform Gaussian smoothing to remove small-scale noise from the raw PM$_{2.5}$ data, since such noise might lead to an incorrect detection of haze events. In our implementation, we employ a 1D discrete kernel with weights [0.1, 0.2, 0.4, 0.2, 0.1].

Step 2 After smoothing the data, we detect the local maxima of the smoothed PM$_{2.5}$ curve. We use a threshold of $t = 115$ that corresponds to the "lightly polluted" air quality, and locate all the local maxima, $\mathscr{P}$, whose PM$_{2.5}$ values are larger than t. To determine candidates for starting and ending times of hazes, we also detect valleys, i.e., local minima, in the PM$_{2.5}$ curve. In particular, we consider only those local minima, $\mathscr{V}$, that are directly connected with a peak whose PM$_{2.5}$ value is larger than t.

Step 3 We identify a haze event as follows:

1. For the first peak $P_0 \in \mathscr{P}$, we locate the immediately preceding local minimum $V_0 \in \mathscr{V}$, which is the starting time of the first haze.
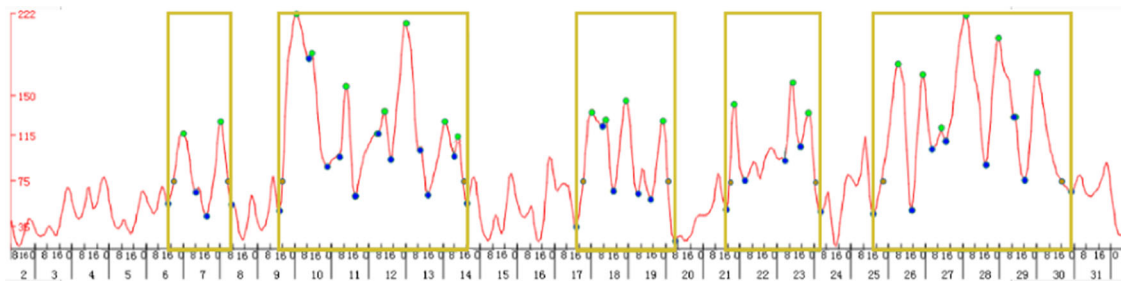

Fig. 4 Detected haze events (highlighted by yellow boxes) using the PM$_{2.5}$ plot. Green and blue dots highlight local maxima and minima of the PM$_{2.5}$ value within a haze event, while orange dots correspond to the times when the PM$_{2.5}$ concentration first (ascending curve) and last (descending curve) reached a value of 75 during the individual hazes
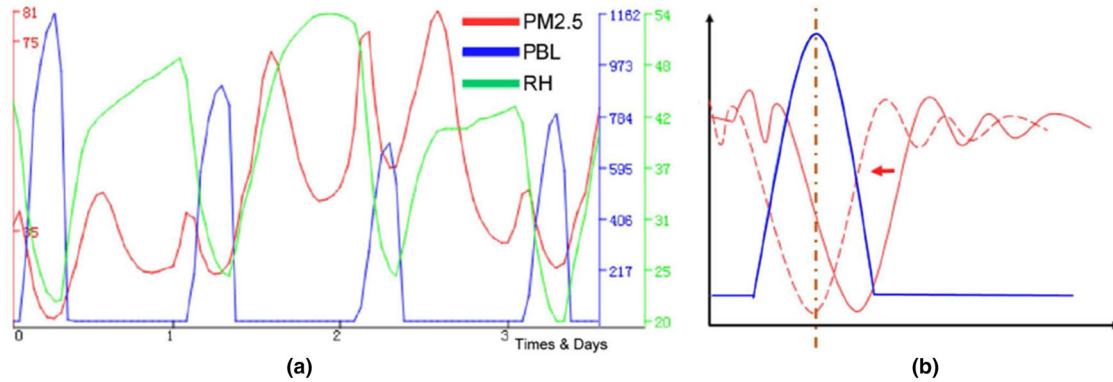
Fig. 5 a The sampled plots of $PM_{2.5}$ (red), RH (green) and PBL (blue). **b** Illustration of phase shifting

2. We now inspect the second peak $P_1 \in \mathscr{P}$. If the time difference between these two peaks is less than 24 h they belong to the same haze event. If there exists a time period that is larger than 24 h between two peaks, but within this period the lasting time where $PM_{2.5}$ is lower than 75 is less than 24 h, we still consider the two peaks belonging to the same haze event.
3. The ending time of this haze period corresponds to the immediate succeeding local minimum of the *last* peak of the period of haze determined by the method described in (2).

We repeat this process for the remaining peaks in $\mathscr{P}$ and the remaining valleys in $\mathscr{V}$ to identify the rest of the haze periods.

5 Correlation among scalar quantities

Similarly to the $PM_{2.5}$ concentration, we summarize all other scalar weather factors, including PBL, RH, temperature and wind strength, into a number of 1D plots with respect to time. We now study their temporal correlations by measuring their corresponding 1D plots. The most challenge problem is that the correlations among $PM_{2.5}$ concentration and other weather factors may experience *delayed* effects in time, i.e., one weather condition is triggered after some time when a correlated condition is satisfied. This usually can be reflected by *phase shifting* in their corresponding 1D plots.

For example, in Fig. 5a, we can observe that the $PM_{2.5}$ (red) and PBL (blue) curves have approximately a negative correlation, i.e., when PBL increases, $PM_{2.5}$ decreases and vice versa. This can be easily revealed by computing their Pearson Product-Moment Correlation Coefficient (PPMCC) (Stigler 1989) using their respective data values during the same period.

However, we observe that the peaks and valleys of these 1D curves do not always coincide, which indicates that their correlation may have some temporal delays. A detailed and formal explanation of the correlation between PBL and $PM_{2.5}$ values will be provided in Sect. 7. In order to capture this delay effect among different 1D curves, we perform a correlation study of the phases between pairs of 1D plots. We refer to this computation as a *phase correlation* computation. We will take the PBL and $PM_{2.5}$ correlation analysis as an example in the following section.

Computation of Phase Shifting We start with computing phase differences between each peak of the PBL plot and the closest valley of the $PM_{2.5}$ curve during a given time period (Fig. 5b). Since the phase differences among multiple periods may not be the same, we propose to leverage the average of phase differences as the delay between the two factors. Specifically, we compute the average of all these phase differences, then shift the $PM_{2.5}$ curve along the X axis based on the average phase difference.

Correlation during hazes As described earlier, hazes in general are associated with multiple weather factors. During a haze event, it is not easy to isolate these factors and solely look at the influence of one factor, say PBL, to the $PM_{2.5}$ concentration. To address this problem, we propose the following solution: First, we estimate the phase shift, s_p, between the $PM_{2.5}$ and PBL curves within a *haze-free* period, determined by the haze event detection algorithm (Sect. 4). Second, we compute the PPMCC coefficient, c, between these two curves with appropriate phase shifting.

Fig. 6 Iso-lines for 2D parameterization

6 Correlation with vector-valued quantities

Different to previously described weather factors, a wind system is often modeled as an unsteady vector field because of its variational directions. Therefore, the correlation between $PM_{2.5}$ and wind cannot be directly obtained using the above approaches for analyzing scalar variables. In this section, we describe how we study the correlation between wind fields and $PM_{2.5}$ concentration using spatial and temporal streamline clustering techniques and particle advection.

We first visualize patterns of the wind field at any given time using streamlines. In our system, we randomly sample seeds within Beijing area and trace out a streamline from each seed using an adaptive Runge-Kutta fourth-order (RK45) integrator in both forward and backward direction. The streamline integration is terminated:

(1) If the velocity value is smaller than some thresholds, i.e., for places with almost no wind; (2) when forming a closed loop (Chen et al. 2007; 3) when reaching the boundary of the data domain; or (4) when reaching the maximum number of integration steps set by the user. Note that we do not terminate the streamline computation even when it is getting really close to another streamline. This is to avoid the generation of many short line segments (see Fig. 6) which may introduce some challenges to the subsequent streamline clustering, e.g., short streamlines may not be classified into the same group with nearby long streamlines.

Based on the preprocessed wind streamlines, we summarize wind patterns by leveraging spatial and temporal streamline clusterings and integrate them into our comparative visual system. We also perform a particle advection scheme to explore origins and destinations of $PM_{2.5}$. Details are shown as below.

6.1 Spatial streamline clustering

We now describe our streamline clustering algorithm inspired from (Yu et al. 2012).

Distance measurements We firstly define two fundamental distance measurements: the distance between two streamlines and the distance between two clusters. Consider two streamlines S_1 consisting of N points, $\mathbf{p}_i$, and S_2 that consists of M points, $\mathbf{q}_j$. The distance between them is measured as follows:

$$
\begin{aligned}
d_{S_1 S_2} &= \min(d_1, d_2) \\
d_1 &= \frac{1}{N} \sum_{i=0}^{N-1} \arg\min_j |p_i - q_j| \\
d_2 &= \frac{1}{M} \sum_{j=0}^{M-1} \arg\min_i |q_j - p_i|.
\end{aligned}
\tag{3}
$$

For each integration point $\mathbf{p}_i$ on S_1, we compute its shortest distance to S_2 by finding the closest point from $\mathbf{q}_j$ to S_2. We then compute the average distance d_1 by summing up these shortest distances for all $\mathbf{p}_i$ and dividing it by N. Similarly, we compute d_2 for S_2. The distance between S_1 and S_2 is then the minimum value of d_1 and d_2. The distance between two clusters C_1 and C_2 is defined as the largest distance between any pair of streamlines S_i and S_j using Eq. (3) where $S_i \in C_1$ and $S_j \in C_2$.

Spatial clustering algorithm With the above distance measurement, a hierarchical streamline clustering can be performed in an iterative process. Initially, each streamline is a cluster. We then procedurally group the closest two clusters into one for each step until only one cluster is left. This gives rise to a hierarchy of streamlines represented by a tree data structure, in which each node is corresponding to a cluster and each leave is an individual streamline.

Please note that in this process, the streamline in a cluster is identified as a representative boundary streamline, when its total distance to the other streamlines in the same cluster is the largest. For the next iteration, another streamline that has the largest distance to the previously identified boundary streamline is found from the rest of the streamlines. The identification of the central streamline is similar, i.e., the total distance of the central streamline to all other streamlines in the cluster is smallest. This simple scheme can greatly facilitate the clustering process. Additionally, we provide an interactive cluster display interface, in which users can control the display of streamline cluster hierarchy. More specifically, clusters with a pairwise distance that is smaller than this threshold will be shown, while clusters in the upper level of the hierarchy will not be displayed, since their distance is larger than the threshold based on the above-given iterative clustering algorithm. When we visualize the selected clusters that satisfy the user-specified threshold, a representative streamline, e.g., the central streamline, for each cluster is shown accordingly.

Figure 7 represents the summarized wind patterns for three sampled times using the above streamline clustering algorithm. In this example 41 streamlines are computed. The summarized wind patterns (highlighted by red curves) did match the expert's expectation. One summarized wind pattern that deserves some attention appeared at 3 AM of January 10. Here, the wind direction tends to *converge* to Beijing area. This coincides with a peak of the PM$_{2.5}$ plot (Fig. 9a) at the same time. A more detailed discussion will be given in Sect. 7.3.

In our experiments, we perform streamline clustering for each sampling time and then generate a comparative visualization to integrate the summarized wind patterns with the plots of PM$_{2.5}$ and the other weather factors. Figure 8 provides an example. The top of this visualization shows a sequence of glyphs that provide the summarized view of the wind patterns at some evenly separated times. The user can inspect the detailed wind system by selecting one of these glyphs.
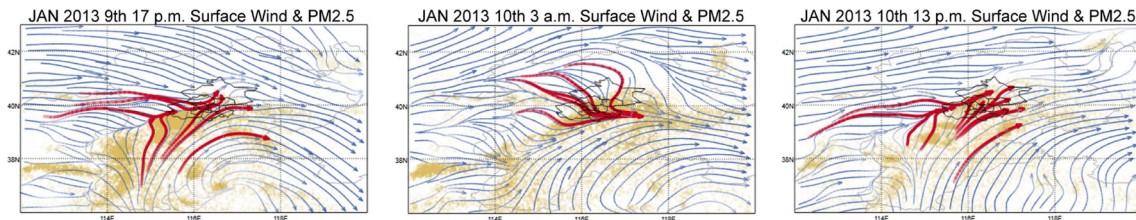

Fig. 7 Examples of summarized wind patterns around Beijing (red curves) for three sampled times in January, 2013 using the proposed spatial streamline clustering technique
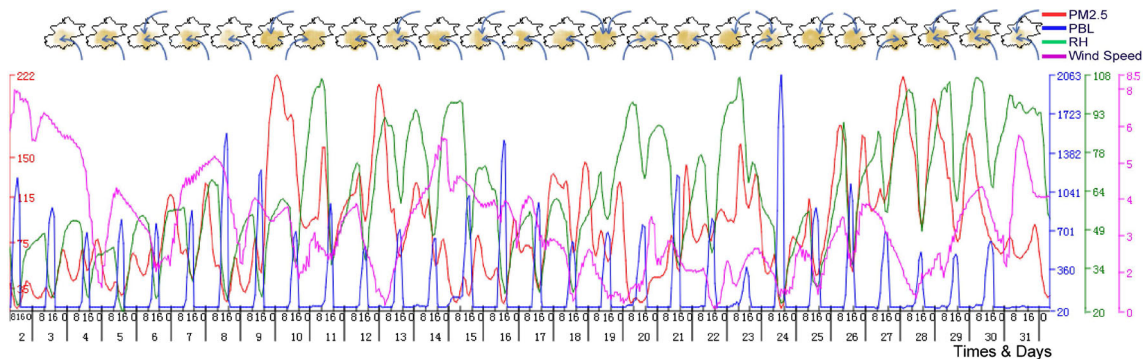

Fig. 8 A comparative visualization of haze related variables within January 2013, where the scalar variables and vector-valued variable are consistently integrated together. On the top we show a sequence of glyphs with the summarized wind patterns in Beijing at the sampled times
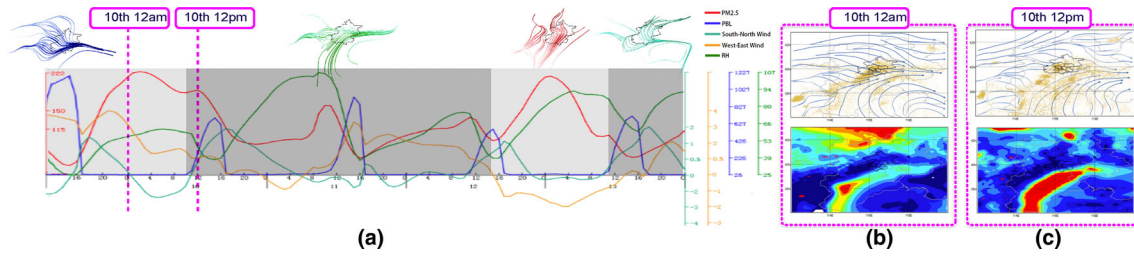
Fig. 9 Exploration of the haze event from January 9, 2013 to January 13, 2013. **a** Temporal streamline clustering where four clusters are obtained. The opacity of the streamlines represents the direction of the wind with darker color corresponding to the downstream and lighter color to the upstream. **b, c** The corresponding PM$_{2.5}$ and wind (above) and relative humidity (bottom) of two time points selected by pink dash lines in **a**

6.2 Temporal streamline clustering

As discussed earlier, during a haze event the influence of PBL to PM$_{2.5}$ concentration is not as obvious as the one that can be observed during a haze-free period (Sect. 7.2). This difference may be caused by the change in the wind field. To study how this change of the wind field influences the PM$_{2.5}$ concentration, we cluster the wind field in time. That is, if during a period of time the wind patterns remain almost the same, we classify them as one group. Through such a temporal clustering, we hope to investigate how changes in the direction and strength of the wind influences PM$_{2.5}$. Specifically, we achieve temporal clustering of the wind field by performing a temporal streamline clustering and group the consecutive instantaneous wind fields into a cluster if they exhibit similar behavior based on their sampled streamlines.

Distance measurements Similar to spatial streamline clustering, there are two distance measures for temporal clustering: (1) the distance between two consecutive instantaneous wind fields and (2) the distance between two temporal clusters. Consider two consecutive instantaneous wind fields, V_a and V_b. Assume N streamlines S_i are computed from V_a, and N streamlines S_j are computed from V_b. The distance between V_a and V_b can then be defined as

$$d(V_a, V_b) = \frac{d_{V_a \to V_b} + d_{V_b \to V_a}}{2} \tag{4}$$

where $d_{V_a \to V_b}$ represents the distance from V_a to V_a, which needs not be identical to $d_{V_b \to V_a}$. The distance V_a to V_b can now be defined as

$$d_{V_a \to V_b} = \arg \max_i \{\arg \min_j d_{S_i S_j}\} \quad \text{with } S_i \in V_a, S_j \in V_b. \tag{5}$$

That is, for each streamline S_i, we identify the streamline from V_b that has the shortest distance to S_i using Eq. (3). The distance from V_a to V_b is then defined as the largest distance from any S_i to V_b. The reason of using the largest distance is that we want to take into account some large local variations. Figure 9 provides an example, where two groups of streamlines (shown in red and blue color) have different starting locations (i.e., the portions of the streamlines drawn in light colors).

After defining the distance between neighboring wind fields at consecutive times, the distance between two temporal clusters C_a and C_b can be simply defined as the distance between the last instantaneous wind field in C_a and the first instantaneous wind field in C_b. This is because C_a and C_b are temporal neighbors, and we assume C_a is located ahead of time to C_b.

Temporal clustering algorithm Having these two temporal distances, we can perform temporal clustering in an iterative fashion. Initially, each instantaneous wind field is a cluster. We then procedurally group the closest two clusters into one at a time until only one cluster is left. Figure 9 provides an example of such a temporal clustering during a haze event that occurred from January 9 toJanuary 13, 2013 (Sect. 7.3).

6.3 Spatial correlation study via particle advection

In this section, we describe how we perform a spatial correlation to help domain experts explore origins and destinations of PM$_{2.5}$ particles during a haze event. In particular, we study particle advection caused by wind fields before and during the specific haze events. In our implementation, we employ a conventional massless particle advection via pathline computation (Eq. 1), instead of recently introduced mass-dependent pathlines
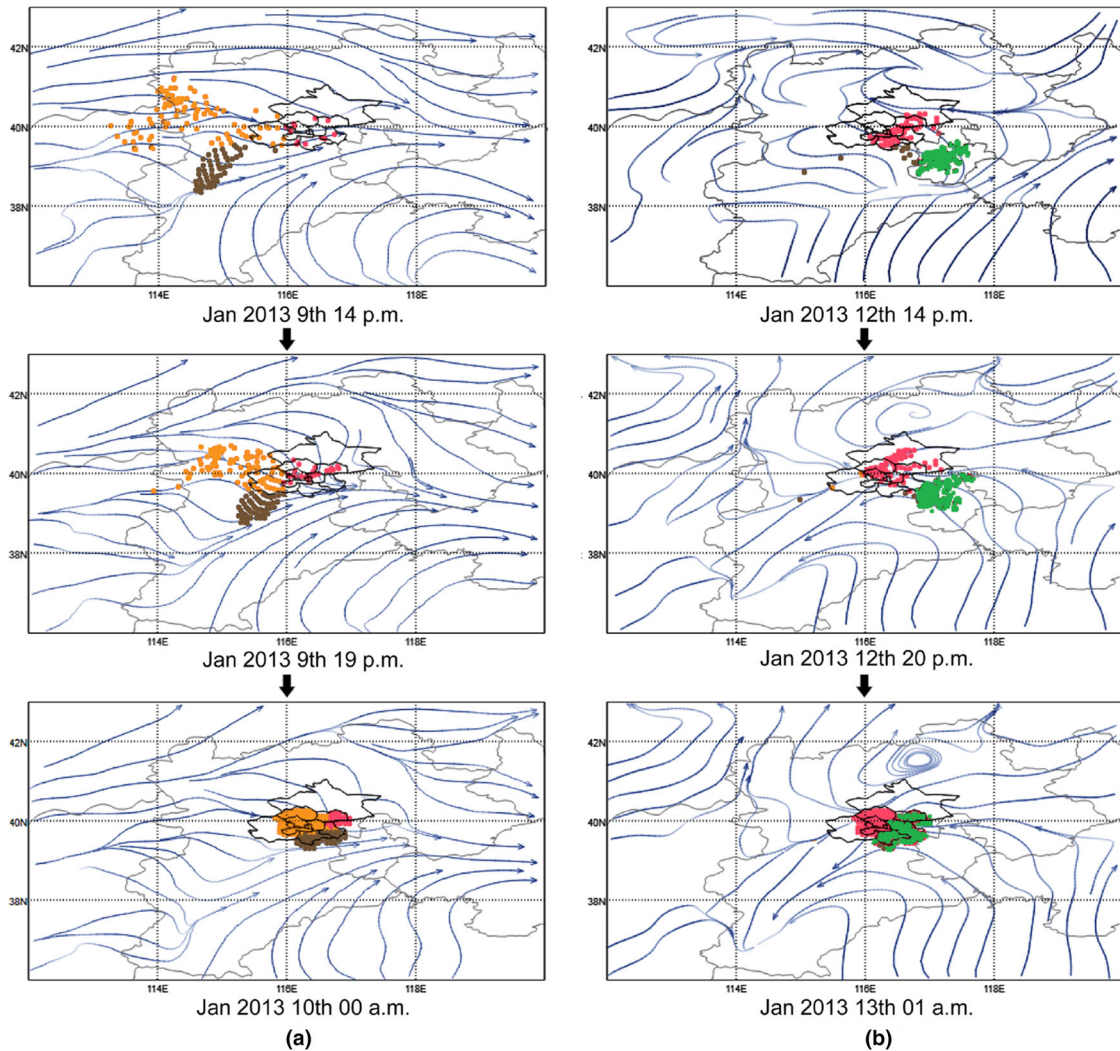
Fig. 10 Two examples of particle backward advection: **a** PM$_{2.5}$ particles moving from Shanxi and Hebei provinces and non-urban regions of Beijing to Beijing; **b** PM$_{2.5}$ particles moving from Tianjin city and the northeast of Hebei province to Beijing

(Günther et al. 2013). This is because the PM$_{2.5}$ particles are much smaller than those discussed in Günther et al. (2013); thus, their mass can be safely neglected.

In our spatial correlation study, we consider two types of particle advections:

- Backward advection to locate the origin of particles when a haze event is starting.
- Forward advection (i.e., in positive time direction) to study the destination of particles when the haze event is ending.

For the backward advection, we proceed as follows: Given the peak time, t_b, and the starting time t_s of the haze event, we advect particles sampled in Beijing area backward starting at t_b and ending at t_s. The density of the particles is determined by the PM$_{2.5}$ concentration distribution in Beijing at t_b. We then inspect which positions those particles can reach, which may indicate the original locations of these particles without worrying about the influence of other weather conditions. Figure 10 provides a number of results of the backward advection for two haze events that occurred in January 9, 2013 (left) and January 13, 2013 (right). Particles with differing colors are sampled at different districts of Beijing. For the haze occurring at January 9, 2013, the visualization indicates that particles were from the west and southwest of Heibei province, while for the haze of January 13 of the same year, the particles were mostly from other regions around Beijing and its neighboring city, Tianjin.

At the same time, we can study where haze particles in Beijing may go during the dissolution of the PM$_{2.5}$ concentration. To achieve that, we advect particles seeded in Beijing forward from t_b to the ending time of the haze at t_e. In combination with backward advection, they render a complete image of where the particles are coming from and where they may go before, in between and after the haze event.

7 Case study

In a number of case studies we evaluated our system. we firstly summarize the main findings based on our system. Then we describe the influence of certain parameters and analyze three haze episodes in detail.

7.1 Key findings

Below we list the key conclusions that we found when experts use our tools.

- When PBL remains in its minimum, the flat portions of the plot PM$_{2.5}$ concentration remain in a high level with some small fluctuation.
- During a haze event, PM$_{2.5}$ concentration increases accordingly with increasing RH; while a decrease in RH may reduce the PM$_{2.5}$ concentration, but not much.
- The V$_{SN}$ may carry pollutant from industrial zone to the urban area of Beijing and increase PM$_{2.5}$ concentration.

7.2 Correlation analysis of climate factors

PBL The planetary boundary layer (PBL) is the lowest layer of the troposphere where wind is influenced by friction. The PBL value, i.e., its thickness, is not constant. It tends to be lower at night and in the cool season, while higher during the day and in the warm season (Haby 2010). Figure 8 provides a comparative visualization of the 1D plots of PM$_{2.5}$ concentration (red) and PBL values (blue). We see that on a typical day, PM$_{2.5}$ concentration starts dropping from 8 AM in the morning and reaches its minimum around 2 PM of the day. It then gradually increases until reaching a maximum around 3 AM of the next day. By overlapping the PM$_{2.5}$ plot and the PBL plot, we see that when PBL values increase, PM$_{2.5}$ values decrease; and when PBL values decrease, PM$_{2.5}$ values increase accordingly. In particular, when PBL remains in its minimum, i.e., the flat portions of the plot PM$_{2.5}$ concentration remain in a high level with some small fluctuation.

This observation can be explained by meteorology (Wang et al. 2014a). Due to the continuous growth of the solar radiation during day time, the ground surface absorbs more heat which leads to an increase in thermal difference between the air and the surface. This difference in turn results in stronger turbulence behavior in troposphere—the lowest layer of the atmosphere, increasing the transportation of the particles from the lower level to higher levels of the atmosphere. Such transportation consequently reduces the concentration of the PM$_{2.5}$ particles. At the same time, the PBL value increases during this process. After 12 PM each day, the heat absorbed by the ground surface decreases, so does the thermal difference between the air and the surface. The turbulence behavior within the troposphere is remitted accordingly. The transportation of PM$_{2.5}$ particles is therefore constrained, leading to the gradual increase in the PM$_{2.5}$ concentration due to the aggregation of PM$_{2.5}$ particles. During this process, the PBL value is decreasing.

Common knowledge in meteorology says that on a *haze-free* day, PM$_{2.5}$ concentration is mainly influenced by the PBL value. However, during a haze event, the high concentration of pollutants reduces the throughput of the solar radiation. The amount of heat absorbed by the ground surface reduces accordingly, leading to small thermal differences between the air and surface. This in turn suppresses the level of turbulence in the troposphere and the corresponding particle transportation.

Wind strength Wind is an important factor that may contribute to the formation of hazes. However, a wind field is a vector field, and therefore, in this section we only focus on its scalar component, i.e., the wind strength—the magnitude of the wind field, and leave the directional component for subsequent processing (Sect. 6). Specifically, we decompose the wind field into its south–north and west–east components, respectively, and study their correlation with hazes. Section 7.3 provide a more detailed insight into this correlation.

Relative humidity Compared to the wind field and the PBL, relative humidity does not play a key role in the variation of the PM$_{2.5}$ concentration. Nonetheless, it is a necessary weather factor when studying haze events. Variations of the humidity may influence a trend of the PM$_{2.5}$ concentration. In general, during a haze event, with increasing humidity PM$_{2.5}$ concentration increases accordingly, while a decrease in humidity may reduce the PM$_{2.5}$ concentration, but not much. A more specific use case on the correlation between RH and PM$_{2.5}$ is provided in the following.

7.3 Comprehensive analysis of hazes in January 2013

Based on our haze event detection results (Figure 4), there existed five noticeable haze events in January 2013. In this section, we focus only on the two main events, i.e., the one occurred between January 9th and 13th (denoted by EVENT #1), and the one from January 26 to 30 (denoted by EVENT #2). According to the above discussion, both the PBL levels and the wind fields play an important role in the evolution of PM$_{2.5}$ concentration. Thus, we mainly concentrate on how these two variables influence the variation of PM$_{2.5}$ during these two periods. As wind field is a vector-valued variable, based on the instruction of an expert, we needed to decompose the wind vector V into two components to facilitate the subsequent analysis. Particularly, we consider two orthogonal directions, i.e., the direction pointing from south to north, $U_\uparrow$, and the direction pointing from west to east, $U_\rightarrow$. The wind vector V can then be decomposed into south–north, V_{SN}, and west–east, V_{WE}, components using $<V, U_\uparrow>$ and $<V, U_\rightarrow>$, respectively. $<,>$ is the inner product of two vectors. The sign of these two components indicates the main wind direction. For example, a negative south–north component indicates that the wind blows from north to south.

Overview of January 2013 Figure 11a shows the plots of PM$_{2.5}$ concentration (red), PBL (blue) and the south–north wind component V_{SN} (light green) of January 2013. Again, we observe a generally negative correlation between PBL and PM$_{2.5}$ values, as shown in Fig. 11b, c where the top image shows the PM$_{2.5}$ concentration and the bottom one shows the PBL. For the relation between V_{SN} and PM$_{2.5}$ we observe that when V_{SN} increases, i.e., the wind mainly blows from south to north, PM$_{2.5}$ tends to increase accordingly. This is because there is an industrial zone south of Beijing. The south–north wind may carry the pollutant with it from this industrial zone to the urban area of Beijing. At the same time, if the V_{SN} is negative, i.e., the wind blowing from north to south, the PM$_{2.5}$ level tends to decrease.

However, we also observe that during some time period the PM$_{2.5}$ concentration was influenced by both PBL and V_{SN}. For instance, from 12 AM to 6 AM of January 10, V_{SN} was negative, while PM$_{2.5}$ concentration reached its peak value. This was caused by the PBL, as at the same time the PBL value was at its minimum. In addition, in an earlier time, i.e., 2 PM–11 PM of January 9, V_{SN} was positive, indicating that the wind blew from south to north, which could carry a large amount of pollutant from the industrial zone to
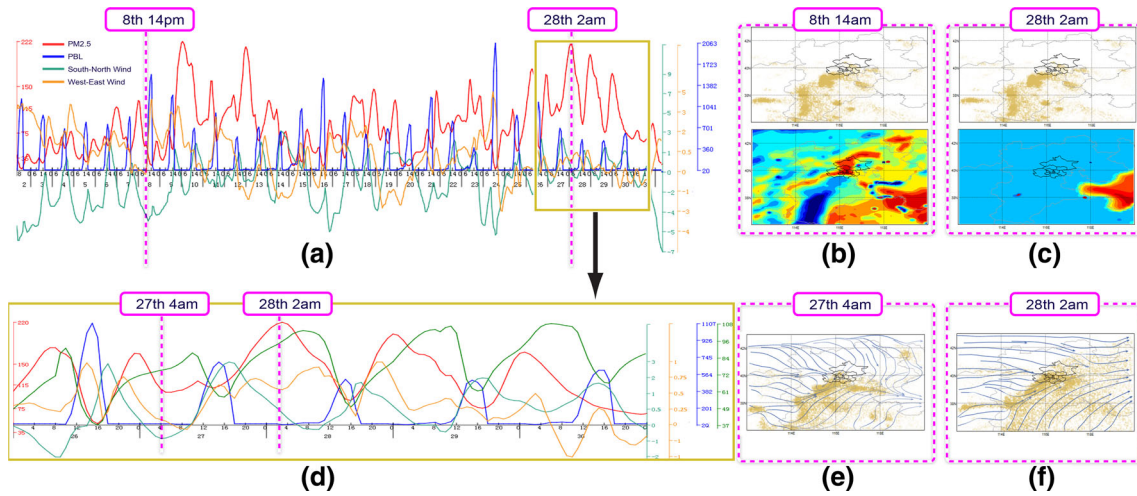

Fig. 11 a Plots of PM$_{2.5}$ concentration (red), PBL (blue), the south–north wind component V_{SN} (dark cyan) and the west–east wind component V_{WE} (orange) on January 2013; **b**, **c** The corresponding PM$_{2.5}$ (above) and PBL (bottom) of two time points indicated by the pink dash lines in (**a**); **d** The magnified view of the haze event highlighted by a yellow box in (**a**); **e**, **f** The corresponding PM$_{2.5}$ and wind field of two time points selected by the pink dash lines in (**d**)

the urban area of Beijing. Combined with the previously described PBL effect, they led to the breakout of the $PM_{2.5}$ particles.

We investigate the correlation between $PM_{2.5}$ concentration and RH which is shown in Fig. 8. Through phase analysis, we found the phase shift between the peaks of RH and the valleys of $PM_{2.5}$ is 0.27, which means the valleys of $PM_{2.5}$ arrive 0.27 h ahead of the peaks of RH in average. After shifting the curve, the correlation coefficient between them is 0.40, which shows a weak positive correlation between RH and $PM_{2.5}$. By separately studying the correlation on haze-free days and hazy episodes, we observe an interesting pattern that the former indicates a negative correlation, while the latter shows a strong positive correlation. With an in-depth analysis of the RH curves during hazy episodes, we find the value of RH is in the range of $60\% - 90\%$ which can help the maintaining of particles. This finding is confirmed by domain experts.

Analysis of EVENT #1 We now analyze the haze occurring from January 9 to 13. Figure 9a shows the plots of $PM_{2.5}$ concentration (red), PBL (blue), V_{SN} (light green) and V_{WE} (orange) within this period. We are focused on the two main peaks during these haze events. The corresponding PBL values at the times of these two peaks were both very low, which did not provide a condition for a dissolution of the $PM_{2.5}$ concentration. Next, we inspected the wind field behavior. During the breakout of the $PM_{2.5}$ at January 10, i.e., during the ascending of the $PM_{2.5}$ plot before reaching the first main peak, both V_{SN} and V_{WE} increased, indicating that the wind blew from southwest to northeast. The $PM_{2.5}$ particles were transported accordingly. At 12 AM of January 10, the wind direction changed from south→north to north→south, indicating the particle aggregation reached its maximum due to the previous transportation process. Coincidentally, the $PM_{2.5}$ concentrations also reached their peak value. After that, the north→south wind started to dominate, the $PM_{2.5}$ value started to decrease as expected. Near 12PM of January 10, the south → north wind started to dominate. Accordingly, $PM_{2.5}$ saw a small increase. But right after 12 PM of the same day, the PBL value gradually increased, which led to a decrease in the $PM_{2.5}$ till dawn of the next day. This can be explained by the influence of PBL to $PM_{2.5}$ as noted in Sect. 7.2. From 12 PM of January 10 to around 4 PM of January 12, the general wind direction is from north to south. Hence, the $PM_{2.5}$ concentration remains in a relatively lower level, although it is still above 75.

Similarly, during the second major breakout which occurred from January 12 to 13, the V_{SN} value played a major role in the variation of the $PM_{2.5}$ concentration.

To learn where the $PM_{2.5}$ particles are coming from during this haze event, we select the first peak point, 12 AM of January 10 to perform a particle backward advection until the starting point of this event, i.e., 2 PM of January 9. Figure 10a shows three snapshots in this process, where we find 28% $PM_{2.5}$ particles came from Shanxi and south of Hebei provinces, 65% $PM_{2.5}$ particles came from the west of Hebei province, and only 7% were from Beijing itself. Meanwhile, we perform the forward particle advection from 2 PM of January 12 to 1 AM of January 13 and find 68% $PM_{2.5}$ particles moved to Tianjin and 23% of them still stayed in Beijing, as illustrated in Fig. 10b. From these numbers, we can see that a large portion of $PM_{2.5}$ particles in Beijing was from Hebei and moved to Tianjin.

Analysis of EVENT #2 The haze event occurring from January 26 to 30 has one obvious difference from the EVENT #1. In EVENT#1, the times when $PM_{2.5}$ reached its peak values typically corresponded to the times when V_{SN} transited from south→north to north→south, as discussed above. From the plots corresponding to EVENT#2 (Fig. 11d), we see that the times when $PM_{2.5}$ reached peaks were in general ahead of the transitioning time of V_{SN} (light green). This is mainly due to the west–east wind, V_{WE} (orange). For instance, let us look at a $PM_{2.5}$ peak occurred from 6 PM of January 27 to 12 PM of January 28. The time when the wind direction switched was about 8 h later than the $PM_{2.5}$ peak. In the meantime, the PBL value was low, creating an ideal environment for the aggregation of haze particles.

Now, let us take the west–east wind into consider for this case. Starting from 4 PM of January 27, the wind direction was mainly from southwest (i.e., both V_{SN} and V_{WE} are positive). At the same time, the $PM_{2.5}$ concentration in the southwest part of Beijing was also high (see Fig. 11e). Therefore, it can be concluded that the southwest wind carried the haze particles to the urban area of Beijing during this period. This resulted in a peak $PM_{2.5}$ concentration value at around 2 AM of January 28 (see Fig. 11f). After that, the wind direction was dominated by the west→east direction, i.e., V_{SN} dropped to near zero, while V_{WE} maintained in a high level. In the meantime, the $PM_{2.5}$ concentration is near zero at the west of Beijing. Consequently, the west→east wind gradually transported the $PM_{2.5}$ particles from Beijing to the outside areas. This explains the delay between the transition time of the south–north wind and the peak time of the $PM_{2.5}$ concentration. All the other $PM_{2.5}$ peaks in EVENT#2 were influenced by the west–east wind in a similar fashion to the above example.

8 Conclusion

In this paper, we propose a visual system to study haze events, including its evolution and correlations with different weather conditions. Our analysis is based on $PM_{2.5}$ concentration, which is employed to measure air quality identify hazes in meteorology. To understand how a haze event is influenced by different weather conditions, we study the correlation between $PM_{2.5}$ concentration and typical weather variables. In particular, we compute the correlation coefficient of $PM_{2.5}$ and other scalar variables, taking into account the delay effect between these variables via a phase correlation computation. We study the correlation of $PM_{2.5}$ with wind fields (unsteady vector fields over time) through a modified spatial streamline clustering method and a novel temporal streamline clustering technique. Additionally, we perform pathline computation to help investigate particle transportation in different temporal phases of haze events. Furthermore, our visual system is evaluated by domain experts on a number of haze events simulated in Beijing area during January 2013.

Future Work The current study focuses on a single layer of the atmosphere and considers only a small subset of the available weather factors. In the future, we plan to extend our system to support studying hazes in multiple layers. The 1D curve representation in our system may become chaotic when weather variables increase. This issue can be improved by introducing hierarchically interactive encodings. Secondly, we will explore other chemical elements in the air, such as PM_{10}, SO_2 and NO_2 concentrations. Existing studies indicate that PM_{10} particles stay mainly local, while $PM_{2.5}$ particles can be transported. It will be interesting to apply our system for analyzing the correlation of PM_{10} concentrations with other weather factors, especially PBL and wind fields. In addition, it is important to combine our method with observation data to verify our haze forecast quality, which we plan to carry out in future works. Finally, we will conduct meteorological experts' reviews in a larger scale to prove its effectiveness and usage for the whole domain area in the future.

References

Berman F, Chien A, Cooper K, Dongarra J, Foster I, Gannon D, Johnsson L, Kennedy K, Kesselman C, Mellor-Crumme J (2001) The grads project: software support for high-level grid application development. Int J High Perform Comput Appl 15(4):327–344

Chen LWA, Chow JC, Doddridge BG, Dickerson RR, Ryan WF, Mueller PK (2003) Analysis of a summertime PM2. 5 and haze episode in the mid-atlantic region. J Air Waste Manag Assoc 53(8):946–956

Chen Y, Cohen JD, Krolik JH (2007) Similarity-guided streamline placement with error evaluation. IEEE Trans Vis Comput Graph 13(6):1448–1455

Ferreira N, Klosowski JT, Scheidegger CE, Silva CT (2013) Vector field k-means: clustering trajectories by fitting multiple vector fields. In: Proceedings of the 15th eurographics conference on visualization, EuroVis '13, pp 201–210

Government C (2012) People's Republic of China ministry of environmental protection standard: technical regulation on ambient air quality index (chinese pdf). http://kjs.mep.gov.cn/hjbhbz/bzwb/dqhjbh/jcgfffbz/201203/W020120410332725219541.pdf/

Günther T, Kuhn A, Kutz B, Theisel H (2013) Mass-dependent integral curves in unsteady vector fields. In: Proceedings of the 15th eurographics conference on visualization. EuroVis '13, vol 32. The Eurographs Association & John Wiley & Sons, Ltd., Chichester, UK, pp 211–220. https://doi.org/10.1111/cgf.12108

Guo H, Yuan X, Huang J, Zhu X (2013) Coupled ensemble flow line advection and analysis. IEEE Trans Vis Comput Graph 19(12):2733–2742

Guo H, Hong F, Shu Q, Zhang J, Huang J, Yuan X (2014) Scalable Lagrangian-based attribute space projection for multivariate unsteady flow data. In: Pacific visualization symposium (PacificVis), 2014 IEEE, pp 33–40. IEEE

Haby J (2010) The planetary boundary layer. http://www.theweatherprediction.com/basic/pbl/

Hankin S, Harrison D, Osborne J, Davison J, O'Brien K (1996) A strategy and a tool, ferret, for closely integrated visualization and analysis. J Vis Comput Animat 7(3):149–157

Hibbard WL (1986) Computer-generated imagery for 4-d meteorological data. Bull Am Meteorol Soc 67(11):1362–1369

Hibbard B, Paul B (1998) vis5d. GNU Public Licenced software, University of Wisconsin Space Science and Engineering Center

Houze RA Jr (2014) Cloud dynamics, vol 104. Academic press, San Diego

http://pharm.vogel.com.cn/2013/0306/news_350750.html

Janicke H, Bottinger M, Mikolajewicz U, Scheuermann G (2009) Visual exploration of climate variability changes using wavelet analysis. IEEE Trans Vis Comput Graph 15(6):1375–1382

Kang CM, Lee HS, Kang BW, Lee SK, Sunwoo Y (2004) Chemical characteristics of acidic gas pollutants and PM2.5 species during hazy episodes in seoul, South Korea. Atmos Environ 38(28):4749–4760

Kehrer J, Ladstadter F, Muigg P, Doleisch H, Steiner A, Hauser H (2008) Hypothesis generation in climate research with interactive visual data exploration. IEEE Trans Vis Comput Graph 14(6):1579–1586

Kehrer J, Muigg P, Doleisch H, Hauser H (2011) Interactive visual analysis of heterogeneous scientific data across an interface. IEEE Trans Vis Comput Graph 17(7):934–946

Ladstadter F, Steiner AK, Lackner BC, Pirscher B, Kirchengast G, Kehrer J, Hauser H, Muigg P, Doleisch H (2010) Exploration of climate data using interactive visualization. J Atmos Ocean Technol 27(4):667–679

Li J, Xiao Z, Zhao HQ, Meng ZP, Zhang K (2016) Visual analytics of smogs in China. J Vis 19(3):461–474

Lu K, Chaudhuri A, Lee TY, Shen HW, Wong PC (2013) Exploring vector fields with distribution-based streamline analysis. In: Proceeding of PacificVis '13: IEEE pacific visualization symposium, Sydney, Australia

Malm WC (1992) Characteristics and origins of haze in the continental United States. Earth Sci Rev 33(1):1–36

McLoughlin T, Laramee RS, Peikert R, Post FH, Chen M (2010) Over two decades of integration-based, geometric flow visualization. Comput Graphics Forum 29(6):1807. https://doi.org/10.1111/j.1467-8659.2010.01650.x

McLoughlin T, Jones MW, Laramee RS, Malki R, Masters I, Hansen CD (2013) Similarity measures for enhancing interactive streamline seeding. IEEE TVCG 19(8):1342–1353

Michalakes J, Dudhia J, Gill D, Henderson T, Klemp J, Skamarock W, Wang W (2004) The weather research and forecast model: software architecture and performance. In: Proceedings of the 11th ECMWF workshop on the use of high performance computing in meteorology, vol 25, p 29

Nocke T, Sterzel T, Bottinger M, Wrobel M (2008) Visualization of climate and climate change data: an overview. In: Digital earth summit on geoinformatics, pp 226–232

Pobitzer A, Lez A, Matkovic K, Hauser H (2012) A statistics-based dimension reduction of the space of path line attributes for interactive visual flow analysis. In: PacificVis, pp 113–120

Potter K, Wilson A, Bremer PT, Williams D, Doutriaux C, Pascucci V, Johhson CR (2009) Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated visus-cdat systems. In: Proceedings of SciDAC 2009, journal of physics: conference series, vol 180, p (published online)

Qu H, Chan WY, Xu A, Chung KL, Lau KH, Guo P (2007) Visual analysis of the air pollution problem in Hong Kong. IEEE Trans Vis Comput Graph 13(6):1408–1415

RenHe Z, Li Q, Zhang R (2014) Meteorological conditions for the persistent severe fog and haze event over eastern China in January 2013. Sci China Earth Sci 57(1):26–35

Sanyal J, Zhang S, Dyer J, Mercer A, Amburn P, Moorhead RJ (2010) Noodles: a tool for visualization of numerical weather model ensemble uncertainty. IEEE Trans Vis Comput Graph 16(6):1421–1430

Schichtel BA, Husar RB, Falke SR, Wilson WE (2001) Haze trends over the United States, 1980–1995. Atmos Environ 35(30):5205–5210

Shi K, Theisel H, Hauser H, Weinkauf T, Matkovic K, Hege HC, Seidel HP (2009) Path line attributes - an information visualization approach to analyzing the dynamic behavior of 3D time-dependent flow fields. In: Hege H-C, Polthier K, Scheuermann G (eds) Topology-based methods in visualization II. Springer, Berlin, pp 75–88

Stigler SM (1989) Francis Galton's account of the invention of correlation. Stat Sci 4:73–79

Sun Y, Zhuang G, Tang A, Wang Y, An Z (2006) Chemical characteristics of PM2. 5 and PM10 in haze-fog episodes in Beijing. Environ Sci Technol 40(10):3148–3155

Sun Y, Jiang Q, Wang Z, Fu P, Li J, Yang T, Yin Y (2014) Investigation of the sources and evolution processes of severe haze pollution in Beijing in January 2013. J Geophys Res Atmos 119(7):4380–4398

Tominski C, Donges JF, Nocke T (2011) Information visualization in climate research. In: 2011 15th international conference on information visualisation (IV), pp 298–305. IEEE

Treinish LA (1994) Severe rainfall events in northwestern peru (visualization of scattered meteorological data). In: IEEE conference on visualization, 1994., Visualization'94, Proceedings., pp 350–354, CP38. IEEE

Wang X, Bi X, Sheng G, Fu J (2006) Chemical composition and sources of PM10 and PM2. 5 aerosols in Guangzhou, China. Environ Monitor Assess 119(1–3):425–439

Wang L, Zhang N, Liu Z, Sun Y, Ji D, Wang Y (2014a) The influence of climate factors, meteorological conditions, and boundary-layer structure on severe haze pollution in the Beijing–Tianjin–Hebei region during January 2013. Adv Meteorol 2014:1–14

Wang Y, Yao L, Wang L, Liu Z, Ji D, Tang G, Zhang J, Sun Y, Hu B, Xin J (2014b) Mechanism for the formation of the January 2013 heavy haze pollution episode over Central and Eastern China. Sci China Earth Sci 57(1):14–25

Wang YH, Fan CR, Zhang J, Niu T, Zhang S, Jiang JR (2015) Forecast verification and visualization based on Gaussian mixture model co-estimation. Comput Graph Forum 34(6):99–110

Wang F, Chen W, Zhao Y, Gu T, Gao S, Bao H (2017) Adaptively exploring population mobility patterns in flow visualization. IEEE Trans Intell Transp Syst 18(8):2250–2259

Yu H, Wang C, Shene CK, Chen JH (2012) Hierarchical streamline bundles. IEEE Trans Vis Comput Graph 18(8):1353–1367

Zhou Z, Ye Z, Liu Y, Liu F, Tao Y, Su W (2017) Visual analytics for spatial clusters for air-quality data. IEEE Comput Graph Appl 37(5):98–105