Spatial-temporal analysis and projection of extreme particulate matter (PM$_{10}$ and PM$_{2.5}$) levels using association rules: A case study of the Jing-Jin-Ji region, China

Shanshan Qin$^{a,b}$, Feng Liu$^{c,d,*}$, Chen Wang$^d$, Yiliao Song$^{c,d}$, Jiansheng Qu$^e$

$^a$ Department Mathematics and Statistics, York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3, Canada
$^b$ MOE Key Laboratory of Western China's Environmental Systems, College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, 730000, China
$^c$ School of Statistics, Dongbei University of Finance and Economics, Dalian, 116025, China
$^d$ School of Mathematics and Statistics, Lanzhou University, Lanzhou, 730000, China
$^e$ Information Center for Global Change Studies, Lanzhou Information Center, Chinese Academy of Sciences, Lanzhou, 730000, China

**Highlights**
- The extreme PM level over Jing-Jin-Ji region of China was investigated.
- Spatial and temporal associations of intercity PM were mined using Apriori algorithm.
- Cross spatial and temporal associations of PM$_{10}$ and PM$_{2.5}$ were also mined.
- Extreme PM frequency and intercity association strength was quantitatively estimated.

**Article Info**

**Abstract**
The Jing-Jin-Ji region of Northern China has experienced serious extreme PM concentrations, which could exert considerable negative impacts on human health. However, only small studies have focused on extreme PM concentrations. Therefore, joint regional PM research and air pollution control has become an urgent issue in this region. To characterize PM pollution, PM$_{10}$ and PM$_{2.5}$ hourly samples were collected from 13 cities in Jing-Jin-Ji region for one year. This study initially analyzed extreme PM data using the Apriori algorithm to mine quantitative association rules in PM spatial and temporal variations and intercity influences. The results indicate that 1) the association rules of intercity PM are distinctive, and do not completely rely on their spatial distributions; 2) extreme PM concentrations frequently occur in southern cities, presenting stronger spatial and temporal associations than in northern cities; 3) the strength of the spatial and temporal associations of intercity PM$_{2.5}$ are more substantial than those of intercity PM$_{10}$.

© 2015 Elsevier Ltd. All rights reserved.

**1. Introduction**

Air pollution issues have received increasing attention in many developing countries due to frequent haze events, which have become a severe threat to people, especially in China (Hu et al., 2013). Historically, serious disasters related to air pollution have been recorded, such as the 1930 Meuse Valley fog (63 people killed), 1948 Donora smog (20 people killed and 7,000 more ill) and the 1952 London smog (4000 more people killed and 100,000 more people ill). These heavy smog events were mainly attributed to accumulated industrial air pollutant emissions and adverse climatic conditions during those periods. As urbanization, industrial development, vehicle usage and other environmentally hazardous processes increase, more fossil fuels are burned, resulting in increased carbon dioxide, nitrogen dioxide, sulfur dioxide, ozone and particulate matter (PM) emissions. Among those emissions, PM (a mixture of solid particles and liquid droplets) is one of the most harmful air pollutants to human health (Hu et al., 2013; Qin et al., 2014; Elbayoumi et al., 2014), especially in extreme PM condition. The World Health Organization (WHO) reported that a short-term PM increase of 10 $\mu$g/m$^3$ can be associated with increased...
coughing, lower respiratory symptoms and hospital admissions involving respiratory problems (WHO, 2000). Exposure to fine particulate matter, PM$_{2.5}$ that is able to highly penetrate the lungs, is strongly correlated with increased rates of mortality, morbidity, respiratory and cardiovascular problems, particularly among children (Elbayoumi et al., 2014; Pope et al., 2009).

To decrease emissions, a detailed PM analysis is necessary for researchers and governmental departments, who seek to further understand the related health effects and formulate effective air pollutant emission control measures to target PM reductions (Yu et al., 2013). Moreover, a dynamic factor analysis (DFA) was developed to identify the interactions among PM and its emission sources (et al., 2014). In recent years, numerous models have been developed to identify the interactions among PM and its emission sources (Chen et al., 2010; Guo; 2011; Winjikui et al., 2015; Ancelet et al., 2014; Hasheminassab et al., 2014; Villalobos et al., 2015; Huertas et al., 2014). Moreover, a dynamic factor analysis (DFA) was developed by extracting the major variables and the associated temporal variations from multivariate time series data (Yu et al., 2013). Kuo et al. (2011) applied the DFA to analyze PM$_{2.5}$ data and spatially identify PM$_{2.5}$ temporal variations in southern Taiwan. Additionally, the nonlinear properties of PM data have been investigated due to the nonlinearities that are inherent in atmospheric systems (Tsontis et al., 1994; Yu et al., 2013). Hu et al. (2013) proposed that examining the spatial and temporal distributions of PM concentrations was useful for reducing air pollution with effective strategies based on the following reasons: spatial analyses can help to identify spatial PM distribution patterns; and temporal analyses of PM time series data are helpful for detecting variation periods and trends. The majority of the pre-existing research focused on original PM data, while few studies analyzed extreme PM levels that would be a most threat to human health.

To take effective measures to control PM levels, it is necessary to understand the extreme spatial and temporal characteristics of PM. Thus, this study focuses on air pollutants collected from Feb. 2014 to Feb. 2015 in the Jing-jin-ji region (encompassing 13 cities), which has always been the most highly air polluted region in China. Because the frequency and severity of the extreme PM values has increased in those cities, joint regional PM research and air pollution control in the Jing-jin-ji region has become an urgent issue. The objective of this study is to examine a comprehensive spatial-temporal analysis and projection for extreme PM issues in the Jing-jin-ji region, initially using an Apriori based association mining technique.

The main contributions of this paper are:

1) an initial examination into the spatial and temporal association rules of intercity PM over Jing-jin-ji using an Apriori algorithm;
2) a quantitative estimate of the extreme PM frequency and maximum projection units;
3) a pioneering attempt to mine the cross spatial temporal associations of PMs (PM$_{2.5}$ and PM$_{10}$) in the Jing-jin-ji region.

This paper is organized as follows: Section 2 introduces the Apriori algorithm based association rules. The study region and data analysis are presented in Section 3. A case study and discussion are presented in Section 4. Finally, we present our conclusions and discussion in Section 5.

2. Apriori algorithm-based association rules

Data mining techniques are typically applied to extract information from large databases and present the data in an easily understandable format (Witten and Frank, 2005). As specific data mining tools, Apriori algorithm-based association rules are adopted to examine the interesting relationships between two or several items that occur synchronously, in the context of information management, decision making, process control and other applications (Guo et al., 2014; Cheng and Li, 2015).

In association rule mining, the database can be stored in the form of transactions, D, and each transaction, T, is composed of non-empty sub-items, I (Soysal, 2015). An association rule is an expression of the form X ⊆ Y. Herein, X ⊆ I, Y ⊆ I. X∩Y = ∅, X and Y are the antecedent and consequent of the rule, respectively (Guzzi et al., 2014). Two metrics are widely used to measure the strength of an association rule, deemed the support and confidence. The support (X → Y) represents the rate of X∪Y contained in D and is given by Eq. (1) (Guo et al., 2014):

\[
support(X \rightarrow Y) = P(X \cup Y)
\]

The support measure evaluates the statistical importance of D. The higher the value, the more important the transaction. Confidence represents the rates of both X and Y contained in D and is given by Eq. (2) (Guo et al., 2014):

\[
confidence(X \rightarrow Y) = \frac{support(X \cup Y)}{support(X)}
\]

where support(X) denotes the rate of X contained in D. The confidence measure evaluates the level of confidence of the association rule.

In addition to these two measures, the lift measure is also applied to identify patterns and trends. The lift (X → Y) is given by Eq. (3) (Soysal, 2015):

\[
lift(X \rightarrow Y) = \frac{confidence(X \rightarrow Y)}{support(X \rightarrow Y)}
\]

The lift measures the dependence of Y on X, which can be interpreted as the degree of lifting confidence between X and Y, from a global existence (support (Y)) to a local association (confidence (X → Y)) (Soysal, 2015). More detailed information of this method can be referenced to (Cheng and Li, 2015; Soysal, 2015).

The association rules are expected to discover the presence of pair conjunctions, which appear in a data set. In time series analyses, intra-transactional association rules are adopted to reveal correlations between multiple time series, but rarely to project time series trends Qian and Shi, 2006). However, association rules do not typically consider temporal relations. To reveal comprehensive and robust relations within extreme PM data from Jing-jin-ji, we performed spatial-temporal association rule mining in this study. Fig. 1 presents a simple process for the proposed rule method.

3. Study region and data analysis

Jing-Jin-Ji is the national capital region of China, and also the largest urbanized region in Northern China. Economic hinterlands exist around Beijing, Tianjin and the Hebei province, along the coast of the Bohai Sea (Jing-Jin-Ji, 2014). This metropolitan region includes 13 major cities, including Baoding, Beijing, Cangzhou, Chengde, Handan, Langfang, Hengshui, Qinhuangdao, Shijiazhuang, Tangshan, Tianjin, Xingtai and Zhangjiakou. The specific location of the Jing-Jin-Ji region, as well as the locations of the major cities, is shown in Fig. 2(1).

This emerging region has experienced a tremendous increase in urbanization in recent decades. It was accompanied by an increase in primary air pollutant emissions and degraded air quality (Feng et al., 2015; Air Pollution in China, 2015). To address the heavy air pollution...
Pollution in Jing-Jin-Ji, China's new air pollution action plan noted that in the Jing-Jin-Ji region, the annual average PM2.5 concentrations should decrease by 26%, 19% and 15%, respectively, in 2017 (CAAC, 2014). In addition, the Beijing, Tianjin and Hebei (also named Ji) governments formulated and circulated their own implementation plans, the Beijing clean air action plan, Tianjin 2012–2020 air pollution control measures plan and the Hebei implementation scheme and action plan for air pollution prevention and control (Yang et al., 2014).

In our study, the hourly concentrations of air pollutants (PM10, PM2.5, NO2, SO2, O3 and CO) were collected in the Jing-Jin-Ji region from Feb. 2014 to Feb. 2015, covering 13 cities and 389 days. Due to the unavailability of electricity, faulty transmission networks and other reasons, missing data must be addressed prior to analysis and modeling. The strategy that was used to assess missing values was the cubic spline Hermite interpolation (CSHI) technique (Qin et al., 2014).

A brief statistical summary of PM10 and PM2.5 is given in Fig. 2(II). Fig. 2(II) presents the overall frequency distributions of PM10 and PM2.5 in the Jing-Jin-Ji region over the study period. The highest frequency of PM10 occurred in the interval from 50–150 μg/m3, with a value of 0.2324, which includes 28,206 data points (the sample size, N, was 121,368). For PM2.5, the highest frequency occurred between 20 and 60 μg/m3, with a value of 0.3328, which includes 40,391 data points. This distribution also shows the relatively extreme PM (PM10 and PM2.5) levels, which are much larger than the 75 μg/m3 and 150 μg/m3 air quality standards for PM2.5 for PM10, respectively. According to the air quality index (AQI) of China, air pollution can be grouped into six levels: excellent, good, lightly polluted, moderately polluted, heavily polluted and severely polluted. Typically, moderately polluted or worse classifications can cause health irritations, meaning that individuals with breathing or heart problems should reduce or avoid outdoor exercise.

The corresponding concentration limits for PM10 and PM2.5 are 250 and 115 μg/m3, respectively. Herein, the PM data are set to be extreme PM if larger than or equal to the specific concentration limits. Fig. 2(III) provides the proportion of extreme PM in 13 cities from the Jing-Jin-Ji region. The pie charts illustrate that Xingtai suffered from the most severe PM pollution, with the highest PM10/PM2.5 frequencies of 33.73% and 43.94%, respectively. Xingtai is followed by another 5 cities that experienced considerably high extreme PM concentrations, being Baoding, Shijiazhuang, Hengshui, Handan and Tangshan, with the extreme PM10/PM2.5 frequencies of 33.73/43.94, 33.39/43.94, 33.39/43.94, 33.39/43.94 and 33.39/43.94, respectively. Conversely, the least amount of PM pollution occurred in Zhangjiakou, accounting for only 2.98% of extreme PM10 and 3.10% of extreme PM2.5, which are better than values from Chengde, Qinhuangdao and Beijing. In the majority of the cities, the PM10 and PM2.5 concentrations followed a similar trend. Taking Beijing, Tianjin and Shijiazhuang as examples, the relationship between PM10 and PM2.5 (see Fig. 2(IV)) exhibits a
Fig. 2. A comprehensive description of the study region. I: The location of the Jing-Jin-Ji region of China; II: The overall distribution of extreme PM (PM$_{10}$ and PM$_{2.5}$) in the Jing-Jin-Ji region; III: The local frequency of extreme PM over 13 cities in the Jing-Jin-Ji region; IV: Scatter plots of PM$_{10}$ versus PM$_{2.5}$ in three representative cities (Beijing, Tianjin and Shijiazhuang); and V: Overall PM boxplots for the Jing-Jin-Ji region for different months, and local PM boxplots for 13 cities from Feb. 2014 to Feb. 2015.
strong linear relationship \((r > 0.88)\). Additionally, Fig. 2(V) illustrates the PM boxplots, which display the spatial and temporal PM data variations. Specifically, Fig. 2(V-1) and Fig. 2(V-2) display the overall PM\(_{10}\) and PM\(_{2.5}\) variations for each month, respectively. In January, April and May, the PM concentrations are relatively small, exhibiting stable variations in the Jing-Jin-Ji region. Comparatively, 3 most severe PM problems occurred in September, October and November, exhibiting high concentrations and variations. From April to November, PM values appeared a gradually rising trend, which means that air quality worsened. Following that period, PM values dropped and remained stable over the next three months. Over the study period (Feb. 2014–Feb. 2015), each city shared similar variations in PM trends (see Fig. 2(V-3) and Fig. 2(V-4)). Outliers are marked with red color. It is indicated that all study cities experienced extreme PM concentrations, which can also be seen in Fig. 2(III).

Because the frequency and severity of the extreme PM values has worsened in those cities, joint regional PM research and air pollution control become an urgent issue in the Jing-Jin-Ji region.

## 4. Results and analysis

In this section, we present the major results of the association rules from the 13 cities in the Jing-Jin-Ji region. Spatial associations can be found in Section 4.1, and spatial-temporal rules are described and discussed in Section 4.2. We set the minimum \(\text{support} = 0.1\) and minimum \(\text{confidence} = 0.7\) in this paper.

### 4.1. Spatial associations among 13 cities in Jing-Jin-Ji region

Table 1 shows the PM\(_{2.5}\) spatial associations from 13 cities in the Jing-Jin-Ji region, which means that association cities (AC, a set that contains certain cities from this region that can influence other cities) were mined and listed in this table. The authors use AC\(_{Beijing}\) to represent Beijing’s AC. It is worth to point that, in Baoding and Xingtai, all of the cities that border them are not always their AC, as shown in Fig. 3 Beijing (bordered by Baoding) and Handan (bordered by Xingtai) are not association cities of Baoding and Xingtai, respectively. In addition, the PM\(_{2.5}\) rules of the 13 cities in the Jing-Jin-Ji region are not totally depended on their spatial positions. For example, Cangzhou, which is near Tianjin, Langfang, Baoding and Hengshui, does not belong to any city’s AC. This situation also occurs in Beijing and Handan. For a clear demonstration of the frequency that city \(A_i\) belongs to other cities’ AC (FeqAC) is defined by Eq. (4).

\[
\text{FeqAC} (\text{City } A_i) = \sum_{j=1}^{N} I_{AC_j}(A_i) 
\]

where \(I_{AC_j}(x)\) is an indicator function (The indicator function for \(AC_j\) if \(x \in AC_j\) equals to 1 if \(x \in AC_j\) and 0 if \(x \in \bar{AC_j}\)). \(r\) and \(\eta\) represent minimum \(\text{support}\) and minimum \(\text{confidence}\) and \(N\) is number of cities.

It is clear that cities whose FeqAC values are 0 have lower \(\text{confidence}\) and \(\text{lift}\) than other cities. This is particularly true of Beijing

| Cities         | Association cities (AC) | Support  | Confidence | Lift   |
|----------------|-------------------------|----------|------------|--------|
| Baoding        | Hengshui                | 0.1758   | 1.0000     | 5.6876 |
| Beijing        | Baoding                 | 0.1584   | 0.7843     | 4.9524 |
|              | Langfang                | 0.1612   | 0.7986     | 4.9524 |
| Cangzhou       | Baoding                 | 0.1758   | 0.9211     | 5.2391 |
| Chendge        | Baoding                 | 0.1758   | 0.7843     | 4.9524 |
|                  | Langfang                | 0.1758   | 0.7986     | 4.9524 |
|                  | Shijiazhuang            | 0.1758   | 0.7986     | 4.9524 |
| Handan         | Baoding                 | 0.1741   | 0.8620     | 4.9524 |
|                  | Langfang                | 0.1758   | 0.8707     | 4.9524 |
|                  | Shijiazhuang            | 0.1758   | 0.9640     | 5.4829 |
|                  | Tangshan                | 0.1758   | 0.9795     | 5.5709 |
|                  | Tianjin                 | 0.1758   | 0.9221     | 5.2391 |
|                  | Xingtai                 | 0.1758   | 1.0000     | 5.6876 |
|                  | Zhangjiaokou            | _        | _          | _      |
and Cangzhou, whose confidence values are below 0.8. Hengshui also did not obtain significant confidence and lift values because FeqAC (Hengshui) is less than Baoding, Langfang, Shijiazhuang, Tangshan, Tianjin and Xingtai, as illustrated in Fig. 4 (The vertical coordinate of Fig. 4 only represents values of FeqAC, confidence and lift). Specifically, the confidence values of these four cities, Beijing, Cangzhou, Handan and Hengshui, are less than 0.9, and their lift values are below 5, which indicates that these cities are significantly different from Baoding, Langfang, Shijiazhuang, Tangshan, Tianjin and Xingtai (when setting confidence = 0.9 and lift = 5, they will be divided into two parts).

In terms of PM10, spatial associations among the 13 cities of the Jing-Jin-Ji region are listed in Table 2. Compared with Table 1, the strengths of the PM10 rules are poorer than those of PM2.5. From the perspective of geographical position, Fig. 5 provides evidence that the rules among the 13 cities are not totally dependent on city position because Xingtai and Handan do not border, Xingtai and Handan, but influence the PM10 concentrations there. It is displayed via the white rectangles, which contain arrows from Xingtai to Baoding and Handan to Shijiazhuang in Fig. 5.

Based on these figures and tables, it is apparent that PM2.5 and PM10 exhibit many differences. When using PM2.5 to search for relationships among cities in the Jing-Jin-Ji region, we found that there is no city associated with them and they do not belong to any cities’ AC. However, when mining the association rules of PM10, there are 8 cities that do not have any relationships with other cities, being Zhangjiakou, Chengde, Qinhuangdao, Beijing, Tianjin, Tangshan, Cangzhou and Hengshui. Among them, Zhangjiakou, Chengde and Qinhuangdao do not have any relationships with other cities regarding rules of PM2.5 or PM10. Considering variations between PM2.5 and PM10, there are two major distinctions: 1) the FeqAC values of PM2.5 and PM10 are significantly different; and 2) the number of elements in each city’s AC decreases from PM2.5 to PM10. Both distinctions might result from the complicative source materials and physical characteristics of PM2.5 and PM10. This distinctions also imply that extreme PM10 occurrences are more independent than extreme PM2.5, and the intercity PM2.5 relationship among the 13 cities is stronger than that of PM10.

### 4.2. Spatial-temporal association rules

After analyzing the spatial relationships among cities in the Jing-Jin-Ji region, spatial-temporal association rules are further mined. These rules can reveal the possible PM movement paths, which indicate how and when city A’s PM influences city B’s PM. In this section, ACBeijing is a set of Beijing association cities with a projection units value of 1.

#### 4.2.1. Spatial-temporal associations of PM2.5

First, the spatial-temporal PM2.5 association rules among these cities will be discussed from a perspective max projection units. AC and the support, confidence and lift of rules are shown in Table 3. In Table 3, Xingtai exhibits the highest value of max projection units, equaling 24. It indicates that extreme PM2.5 events in this city can be deduced according to its AC’s condition within 24 h with a confidence of 0.7593. There are three cities whose max projection is 0, which indicates they do not have any temporal relationships with other cities. From a geographic perspective, those three cities are located in the northern portion of the region, reflecting that extreme PM2.5 events typically occur in the southern portion of the region. Given that the max projection units of Baoding, Handan, Shijiazhuang and Xingtai are over 10, and both Beijing and Tianjin

### Table 2

| Cities               | AC          | Support     | Confidence | Lift          |
|----------------------|-------------|-------------|------------|---------------|
| Baoding              | Langfang    | Xingtai     | 0.2022     | 0.9099        | 4.4995        |
| Shijiazhuang         | Handan      | Xingtai     | 0.3219     | 1.0000        | 3.1062        |
| Xingtai              | Hengshui    | Shijiazhuang| 0.1903     | 0.7809        | 4.1037        |

Note: AC, Support, Confidence and Lift of Beijing, Cangzhou, Chengde, Handan, Hengshui, Langfang, Qinhuangdao, Tangshan, Tianjin and Zhangjiakou do not exist.

Fig. 4. Each city’s FeqAC, confidence and lift in spatial association rules of PM2.5.

Fig. 5. Spatial association rules of PM10.
are core cities, thereby those 6 cities will be used to demonstrate their specific association rules in this region.

In Table 4, from the perspective of $A_{\text{Beijing, i}}$ ($i = 0, 1, \ldots, 5$), there are two groups, {Baoding, Langfang, Shijiazhuang, Tangshan, Tianjin, Xingtai} and {Baoding, Handan, Hengshui, Langfang, Shijiazhuang, Xingtai}. The former’s projection units values range from 0 to 2, while the latter’s range from 3 to 5. From Table 4, it is apparent that Tianjin and Tangshan do not influence Beijing’s PM$_{2.5}$ when the projection units value is greater than 2, while Handan and Hengshui will affect Beijing within three hours. It reflects that Beijing will not be influenced by the high PM$_{2.5}$ concentration that occurs three hours ago in those cities located to the east of Beijing. However, the extreme PM$_{2.5}$ in the cities to the south of Beijing will influence Beijing’s PM$_{2.5}$ significantly, implying that those southern cities’ high PM$_{2.5}$ occurrences would play a key predictor to Beijing’s extreme PM$_{2.5}$, especially within 3–5 h.

Tianjin’s spatial-temporal rules are listed in Table 5. The projection units of Tianjin is larger than that of Beijing, which suggests that we can deduce Tianjin’s high PM$_{2.5}$ events longer than Beijing’s. In this table, $A_{\text{Tianjin, i}}$ ($i = 0, 1, \ldots, 7$) is divided into three groups, when $i$ is equal to 2, 5, 7, respectively. Through observing the geographic positions and projection units of these three groups, a similar conclusion can be drawn that cities located to the south of Tianjin will influence Tianjin’s pollution situation in 6–7 h if high PM$_{2.5}$ events occur currently.

Detailed information for other cities is shown in Fig. 6. It illustrates the mining results. In this figure, the horizontal axis and vertical axis of each subgraph represent projection hours and city_NUM, which is shown at the bottom of the figure, respectively. Each point in a subgraph represents an association city. This means that the AC for each projection hours, $K$, will be obtained through transforming the city_NUMs of points to city names using the information at the bottom of the figure. Specifically, [1 5 7 10 11 12] means {Baoding, Langfang, Tangshan, Tianjin, Xingtai}.
southern cities in $A_{C_{Baoding, i}} \ (i = 5, 6, 7, 8)$ increases. Overall, when the projection units value is greater than 14, Tangshan and Tianjin, the northern cities in the Jing-Jin-Ji region, do not affect Baoding. However, Handan and Hengshui, both southern cities, give rise to influence Baoding’s PM$_{2.5}$. This means that southern cities have an increasing effect when the projection units increase in Baoding. Fig. 6(d) illustrates the spatial-temporal rules in Xingtai with the highest max projection units among 13 the cities in the region. It is different from other cities because there are five main groups in Xingtai, which are represented by blue, red and green circles and red and blue rectangles, whose number of elements is 2, respectively. There are two major changes in Fig. 6(d). Both changes suggest that the number of elements in $A_{C_{Xingtai, i}}$ decreases, while the ratio of southern cities in $A_{C_{Xingtai, i}}$ increases. Thus, the main spatial-temporal PM$_{2.5}$ rules have been demonstrated, and the next step is to illustrate the PM$_{10}$ rules.

4.2.2 Spatial-temporal associations of PM$_{10}$

In Table 6, Hengshui exhibits the highest max projection units, equaling 6, which is less than the highest max projection units in Table 3. The max projection units values greater than 0 occur in only four cities, which suggests that the intercity PM$_{10}$ relationship

![Fig. 6. The spatial-temporal association rules for PM$_{2.5}$ in four cities: Shijiazhuang, Handan, Baoding, Xingtai.](image-url)
among the 13 cities in this area is poorer than that of PM$_{2.5}$. From this table, Xingtai will experience a similar situation within 5 h if Baoding and Langfang suffer extreme PM$_{10}$. For the other two major cities, Beijing and Tianjin, the results indicate that their PM$_{10}$ pollution does not correlate with other cities in the Jing-Jin-Ji region.

For Baoding, there are two AC groups, which are (Langfang, Xingtai) when the projection units values range from 0 to 2, and (Hengshui, Xingtai) when the projection units value is 3. This indicates that southern cities have a greater effect on Baoding as projection units increase because Hengshui is located further south than Langfang is. A similar situation occurs in Shijiazhuang because there are also two groups of AC, one of which, (Handan, Xingtai), appears when the projection units values range from 0 to 2. The other, (Hengshui, Xingtai), appears only when the projection units value is 3. Both Hengshui and Xingtai have only one AC group. The former’s group is (Shijiazhuang, Xingtai) and latter’s is (Hengshui, Shijiazhuang).

### 4.2.3. Cross spatial-temporal associations of PM$_{2.5}$ and PM$_{10}$

To this point, we have discussed the spatial-temporal PM$_{2.5}$ and PM$_{10}$ associations. However, it remains unknown if a specific city’s PM$_{10}$ will be affected by other cities’ extreme PM$_{2.5}$, and vice versa. To answer this question, spatial-temporal association rules are mined based on cross-PMs (see Tables 7 and 8). For Table 7, it is clear that there are only two cities whose PM$_{10}$ values exhibit relations with the PM$_{2.5}$ values of other cities. This indicates that the intercity PM$_{10}$ relationships are stronger than the cross-PMs (PM$_{2.5}$ → PM$_{10}$) relations. However, for a single city, the max projection units values of Baoding and Xingtai in Table 7 are larger than those in Table 6, which means that PM$_{10}$ influences PM$_{2.5}$ less strongly than PM$_{2.5}$ does in both cities. In Fig. 7, the detailed cross spatial-temporal associations (PM$_{2.5}$ → PM$_{10}$) of the Baoding and Xingtai are illustrated. From this figure, it is clear that southern cities have a greater effect on Baoding and Xingtai as the projection units value increases.

The results from another cross case are shown in Table 8. In this table, the max projection units value of Shijiazhuang is equal to that of Table 3, which means that the PM$_{10}$ of Hengshui and Xingtai can influence the succeeding 10 h of PM$_{2.5}$ values in Shijiazhuang. However, the max projection units values for other cities are less than those in Table 3. So, the cross spatial-temporal rules of Shijiazhuang will be discussed specifically.

Table 9 shows the cross spatial-temporal association rules (PM$_{10}$ → PM$_{2.5}$) in Shijiazhuang. Clearly, there are three groups in the column “AC”, being (Baoding, Hengshui, Xingtai), (Handan, Xingtai) and (Hengshui, Xingtai), respectively. From a geographic perspective, southern cities increasingly influence Shijiazhuang as the projection units increases because Baoding does not belong to AC after the projection units value reaches 4.

### 5. Conclusions and discussion

This study presents a spatial-temporal analysis and projection of extreme PM concentrations around the Jing-Jin-Ji region of China during the period from Feb. 2014 to Feb. 2015. The Apriori algorithm-based association rules mining technique is used to discover the spatial-temporal relations between a target city and other potential reference cities. The findings and their explanations are summarized as follows:

1. The intercity PM associations do not completely rely on spatial distributions, which means that there are no associations among several cities that border one another, and vice versa. For example, Cangzhou is surrounded by Tianjin, Langfang, Baoding and Hengshui, but not associated with...
Fig. 7. The cross spatial-temporal associations in two cities (Baoding and Xingtai) — Part I (PM$_{2.5}$ / PM$_{10}$).

Table 9
Cross spatial-temporal association rules (PM$_{10}$ / PM$_{2.5}$) in Shijiazhuang.

| Projection units | AC        | Support | Confidence | Lift  |
|-------------------|-----------|---------|------------|-------|
| 0                 | Baoding   | Hengshui| Xingtai    | 0.2007| 0.9013  | 4.4903 |
| 1                 | Baoding   | Hengshui| Xingtai    | 0.1994| 0.8970  | 4.4996 |
| 2                 | Baoding   | Hengshui| Xingtai    | 0.1972| 0.8903  | 4.5148 |
| 3                 | Baoding   | Hengshui| Xingtai    | 0.1913| 0.8709  | 4.5519 |
| 4                 | Baoding   | Hengshui| Xingtai    | 0.1846| 0.8489  | 4.6000 |
| 5                 | Handan    | Xingtai |            | 0.2978| 1.0000  | 3.3583 |
| 6                 | Handan    | Xingtai |            | 0.2894| 1.0000  | 3.4557 |
| 7                 | Handan    | Xingtai |            | 0.2814| 1.0000  | 3.5540 |
| 8                 | Hengshui  | Xingtai |            | 0.2729| 1.0000  | 3.6641 |
| 9                 | Hengshui  | Xingtai |            | 0.2665| 1.0000  | 3.7520 |
| 10                | Hengshui  | Xingtai |            | 0.2606| 0.9894  | 3.7974 |
them. This result also agrees with that in Beijing and Handan. The low support and confidence values in Beijing imply that PM concentrations therein are poorly associated with other reference cities. It might be explained by that the PM pollution in Beijing mainly comes from intra-city emissions, such as fossil fuel combustion, which is probably due to the significant vehicle increasing in recent years. This rising trend is expected to continue.

(2) Over the Jing-Jin-Ji region, extreme PM events frequently occur in the southern cities, and the strength of the spatial and temporal association PM rules is higher for southern cities than northern cities. In addition, PM in southern cities has a greater influence on northern cities as projection units values increase. This result can be explained by the topography and wind direction in this region. From the perspective of topography, the Taihang Mountains are located to the west of the Jing-Jin-Ji region. Therefore, strong winds from the west are extremely rare, which was statistically estimated by the historical weather data (Lishi Tianqi, 2015). This indicates that PM can move in three possible directions, being north to south, south to north and east (the sea) to west (the mountains) if wind is responsible for PM transportation. For the recorded data (Lishi Tianqi, 2015), the average blowing time from the south in southern cities is longer than from the north in northern cities. This indicates that PM from southern cities has a high probability to be transported northward by the wind. This conclusion from topography and wind direction agrees with the result that PMs of southern cities influences that of northern cities more. Therefore, a substantial reduction in PM emissions in the southern cities of the Jing-Jin-Ji region is of critical significance for reducing PM concentrations and improving the air quality across this region.

(3) The spatial and temporal intercity PM$_{2.5}$ associations are much stronger than PM$_{10}$. This is because PM$_{2.5}$ is suspended in the air for a longer period of time and is easily transported between cities. PM$_{2.5}$ in Jing-Jin-Ji can be attributed to multiple sources and is mainly dispersion-governed. PM$_{10}$ is likely of local origin and more affected by traffic. It was estimated that the contribution of traffic to coarse mode urban ambient concentrations was larger on weekdays than that on weekends, due to higher traffic flows (Barmpadimios et al., 2011). Gugamsetty et al. (2012) suggested that soil dust and vehicle emissions were the major sources of PM$_{10}$, and the dust caused by traffic would inevitably increase PM$_{10}$ concentrations. The higher intercity PM$_{2.5}$ associations also indicate that spatial and temporal PM$_{2.5}$ projections can be easily and reliably conducted. When extreme PM$_{2.5}$ events occur in certain reference cities, the target city will suffer extreme PM$_{2.5}$ at a probability of greater than 0.7. However, the same cannot be said for PM$_{10}$, due to poor intercity PM$_{10}$ associations.

(4) The maximum projection units are also estimated, through which an extreme PM incidence can be predicted 1-maximum projection in advance. In addition, as the projection steps increase, the association cities for a target city will change, and their association strengths rarely decline.

We believe this is an initial research to identify the spatial and temporal association rules of extreme PMS for the Jing-Jin-Ji region, an area but severely polluted area around the capital of China. This investigation was performed using an Apriori algorithm, which was demonstrated to be an effective method for extreme PM data mining. The mining results agreed with the real world situation, suggesting that they were reasonable and reliable. These results may reveal helpful information for researchers and policy makers regarding PM pollution. The mining results can be applied to formulate effective strategies for reducing extreme PM incidents and improving air quality in the region. In addition, this research provides each city’s max prediction steps. These values are beneficial for building PM$_{2.5}$ and PM$_{10}$ warning systems, which can reduce the harmful effects of pollution and enable relevant parties to take the necessary precautions.

Acknowledgments

The work was funded by the National Natural Science Foundation of China (Grant No. 71171102) and the China Scholarship Council.

References

Air Pollution in China. 2015. Air Pollution in the Beijing-Tianjin-Hebei (BTH) Region. http://www.journals.elsevier.com/atmospheric-environment/call-for-papers-air-pollution-in-china-air-pollution-in-the-beijing-tianjin/

Ancelet, T., Day, P.K., Trompetter, W.J., Markowitz, A., Weatherburn, D.C., 2014. Sources and transport of particulate matter on an hourly time-scale during the winter in a New Zealand urban valley. Urban Clim. 10, 644–655.

Barmpadimios, I., Nuber, M., Oederholz, D.C., Keller, J., Aksoyoglu, S., Hugelin, C., Prévet, A.S.H., 2011. The weekly cycle of ambient concentrations and traffic emissions of coarse (PM 10–PM 2.5) atmospheric particles. Atmos. Environ. 45 (27), 4580–4590.

CAAC, 2014. Can Beijing, Tianjin and Hebei Achieve their PM$_{2.5}$ Target by 2017. Scientific Assessment Showed Enhanced Policy Would Enable Jing-Jin-Ji to Achieve 2017 Targets. http://en.cleanairchina.org/product/6659.html.

Chen, L.W.A., Watson, J.G., Chow, J.C., DuBois, D.W., Herschberger, L., 2010. Chemical mass balance source apportionment for combined PM 2.5 measurements from US non-urban and urban long-term networks. Atmos. Environ. 44 (38), 4988–4998.

Cheng, Y., Li, Q., 2015. GA-based multi-level association rule mining approach for defect analysis in the construction industry. Automation Constr. 51, 78–91.

Elbakroun, M., Ramli, N.A., Yusof, N.J.F.M., Yahaya, A.S.R., Al Madhoun, W., Ul-Saufie, A.Z., 2014. Multivariate methods for indoor PM 10 and PM 2.5 modelling in naturally ventilated schools buildings. Atmos. Environ. 94, 11–21.

Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J., 2015. Artificial neural networks forecasting of PM 2.5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmos. Environ. 107, 118–128.

Gugamsetty, B., Wei, H., Liu, C.N., Awashti, A., Hsu, S.C., Tsai, C.J., Chen, C.F., 2012. Source characterization and apportionment of PM$_{10}$, PM$_{2.5}$ and PM$_{1}$ by using positive matrix factorization. Aerosol Air Qual. Res. 12, 476–491.

Guo, H., 2011. Source apportionment of volatile organic compounds in Hong Kong homes. Build. Environ. 46 (11), 2280–2286.

Guo, Z., Chi, D., Wu, J., Zhang, W., 2014. A new wind speed forecasting strategy based on the chaotic time series modelling technique and the Apriori algorithm. Energy Convers. Manag. 84, 140–151.

Guazzi, P.H., Milano, M., Cannataro, M., 2014. Mining association rules from gene ontology and protein networks: promises and challenges. Procedia Comput. Sci. 29, 1970–1980.

Hasheminassab, S., Daher, N., Ostro, B.D., Soutas, C., 2014. Long-term source apportionment of ambient fine particulate matter (PM 2.5) in the Los Angeles basin: a focus on emissions reduction from vehicular sources. Environ. Pollut. 193, 54–64.

Hu, M., Jia, L., Wang, J., Pan, Y., 2013. Spatial and temporal characteristics of particulate matter in Beijing, China using the empirical mode decomposition method. Sci. Total Environ. 458, 70–86.

Huertas, J.I., Huertas, M.E., Cervantes, G., Díaz, J., 2014. Assessment of the natural sources of particulate matter on the opencast mines air quality. Sci. Total Environ. 493, 1047–1055.

Jing-Jin-Ji (2014). http://www.jingjinji.com/wiki/Jing-Jin-Ji.

Kuo, Y.M., Wang, S.W., Jiang, C.S., Yeh, N., Yu, H.L., 2011. Identifying the factors influencing PM 2.5 in southern Taiwan using dynamic factor analysis. Atmos. Environ. 45 (39), 7276–7285.

Lishi Tianqi (2015). http://lishi.tianqi.com/.

Pope III, C.A., Ezzati, M., Dockery, D.W., 2009. Fine-particulate air pollution and life expectancy in the United States. N. Engl. J. Med. 360 (4), 376–386.

Qin, L.X., Shi, Z.Z., 2006. Efficiently mining association rules from time series. Int. J. Inf. Technol. 12 (4), 30–38.

Qin, S., Liu, F., Wang, J., Sun, B., 2014. Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models. Atmos. Environ. 98, 665–673.

Soyal, O.M., 2015. Association rule mining with mostly associated sequential patterns. Expert Syst. Appl. 42 (5), 2582–2592.

Tsonis, Reviewed by, Anastasios, A., Ditto, William L., 1994. Chaos: from theory to application. Am. J. Phys. 62, 2.

Villalobos, A.M., Barraza, F., Jorquera, H., Schauer, J.J., 2015. Chemical speciation and
source apportionment of fine particulate matter in Santiago, Chile, 2013. Sci. Total Environ. 512, 133–142.
WHO, 2000. Air Quality Guidelines for Europe World Health Organization. WHO, Copenhagen.
Winijkul, E., Yan, F., Lu, Z., Streets, D.G., Bond, T.C., Zhao, Y., 2015. Size-resolved global emission inventory of primary particulate matter from energy-related combustion sources. Atmos. Environ. 107, 137–147.

Witten, I.H., Frank, E., 2005. Data mining: practical machine learning tools and techniques. Morgan Kaufmann.
Yang, S., Chen, B., Fath, B., 2014. Trans-boundary Total Suspended Particulate Matter (TSPM) in Urban Ecosystems. Ecological Modelling.
Yu, H.L., Lin, Y.C., Sivakumar, B., Kuo, Y.M., 2013. A study of the temporal dynamics of ambient particulate matter using stochastic and chaotic techniques. Atmos. Environ. 69, 37–45.