Analysis of insoluble particles in hailstones in China

Haifan Zhang[1], Xiangyu Lin[1], Qinghong Zhang[1*], Kai Bi[2*], Chan-Pang Ng[1], Yangze Ren[1], Huiwen Xue[1], Li Chen[3], Zhuolin Chang[4]

[1]Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871, China
[2]Field Experiment Base of Cloud and Precipitation Research in North China, China Meteorological Administration, Beijing 101200, China
[3]Electron Microscopy Laboratory, Peking University, Beijing 100871, China
[4]Key Laboratory for Meteorological Disaster Monitoring and Early Warning and Risk Management of Characteristic Agriculture in Arid Regions, China Meteorological Administration, Yinchuan 750002, China

**Corresponding author*: Qinghong Zhang (qzhang@pku.edu.cn); Kai Bi (bikai_picard@vip.sina.com)

Abstract. Insoluble particles ~~affect~~influence weather and climate ~~indirectly~~ by heterogeneous freezing process. Current weather and climate models ~~have large uncertainty~~face considerable uncertainties in freezing processes simulation due to ~~little~~limited information regarding species and number concentration of heterogeneous ice-nucleating particles, ~~mainly~~particularly insoluble particles. Here, for the first time, size distribution and species of insoluble particles are analyzed in 30 shells of 12 hailstones ~~in~~collected from China, using scanning electron microscopy and energy dispersive X-ray spectrometry. ~~Total~~A total of 289,461 insoluble particles ~~a~~were detected and ~~grouped~~identified into 3 species: organics, dust, and bioprotein~~ by~~, utilizing machine learning methods. The size distribution of insoluble particles of each species var~~y~~ies greatly ~~in~~among the different hailsto~~rm~~nes but little in their shells. Further, classic size distribution ~~modes~~ of organics and dust ~~were performed as~~followed logarithmic normal distributions, which ~~may~~could potentially be adapted in future weather and climate models~~ though uncertainty still exists.~~, despite the existence of uncertainties. Our finding suggests ~~that physical properties of aerosols~~the aerosol species and number concentration variance in different storms should be considered in model simulation ~~on~~of the ice freezing process.

1 Introduction

Insoluble particles, acting as main heterogeneous ice-nucleating particles in the atmosphere(Lamb and Verlinde, 2011), ~~may indirectly impact~~influence precipitation formation and radiative forcing (Hoose and Möhler, 2012; DeMott et al., 2015), and further impact weather and climate (Vergara-Temprado et al., 2018). Temperature and vapor supersaturation are used to calculate the number concentration of ice crystal particles in microphysics parameterization rather than considering the physical properties of ice-nucleating particles in weather and climate models (DeMott et al., 2010). ~~Only few~~Few models ~~calculate~~used the freezing parameterization, which establishes a direct connection between the number concentration of ice-nucleating particles and the number concentration of ice crystals. The absence of description regarding the number concentration of ice-nucleating particles in ~~clouds, that leads~~models can result in an incorrect estimation of ice crystals and lead to ~~a misestimation about number concentration of ice particles and large errors in simulation~~significant bias in radiative simulations (Vergara-Temprado et al., 2018).

An improved description ~~for~~of the number concentrations of ice-nucleating particles is needed~~, while obstructed by a lack of complete microphysical observation in clouds about ice-nucleating particles~~ (DeMott et al., 2010)~~. Measurements of the number concentration and species of ice-nucleating particles, mainly insoluble particles(Lamb and Verlinde, 2011), were conducted by an airborne equipment or laboratory instrument with air parcels, to understand the process of ice nucleation in clouds(DeMott et al., 2010; Prenni et al., 2009; Hoose et al., 2010; Rogers et al., 2001). Most field projects sampled air parcels in anvils of convective clouds, cirrus and winter mixed-phase stratiform clouds, keeping airborne equipment in good working condition. However, few projects sampled air parcels through cores in convection. Thus, current observation is insufficient for~~

describing the whole convective cloud, especially the deep convection in severe storms. Absence about microphysical observations of ice-nucleating particles within severe storms leads to uncertainty in understanding cold cloud process, e.g., hailstone formation, while obstructed by a lack of complete microphysical observation in clouds about ice-nucleating particles. There are two ways to sample ice-nucleating particles: The first involves an airborne instrument, named continuous flow thermal gradient diffusion chamber (Rogers et al., 2001; Prenni et al., 2009; DeMott et al., 2010). The second is done in the laboratory, where scientists conduct freezing experiments (Hoose and Möhler, 2012). In most cases, it is necessary for an aircraft to collect air parcels for measurement of the physical properties of ice-nucleating particles in the air. However, former field projects sampled air parcels in anvils of convective clouds, cirrus and winter mixed-phase stratiform clouds. No flight report or article has reported that they sampled air parcels through cores in deep convection. This phenomenon is consistent with consideration for flight security. Thus, current observation is insufficient for describing the whole convective cloud, especially the deep convection in severe storms. Absence about microphysical observations of ice-nucleating particles within severe storms leads to uncertainty in understanding cold cloud process.

Hailstones, as a product of deep convective clouds, serves as a carrier of information within these clouds. Recently,(Li et al., 2020).

Recently, detection for soluble ions along with isotopic analysis of a huge hailstone revealed an up-and-down hailstone growth trajectory, which demonstrated that the different shells were formed at different heights (Li et al., 2020). Further analysis revealed large diversity in number concentration of soluble ions among hailstones from different hailstorms (Li et al., 2018). Further, the detection of soluble ions along with isotopic analysis of a huge hailstone revealed an up-and-down hailstone growth trajectory, which demonstrated that the different shells were formed at different heights (Li et al., 2020). These studies have proved aerosol information in convective cloud may be recorded in soluble particles within hailstones(Li et al., 2020, 2018; Knight, 1981; Jouzel et al., 1975).These studies have proved that aerosol information in convective cloud may be recorded in soluble particles within hailstones (Li et al., 2018, 2020). Similarly, insoluble particles in hailstones can also record aerosol information in severe storms.

Former studies showed that species and number concentration of insoluble particles in hailstones (Vali, 1968; Rosinski, 1966; Michaud et al., 2014) would influence heterogeneous nucleation process (Hoose and Möhler, 2012) and further hailstone formation (Knight, 1981). Information on the species of insoluble particles can determine the freezing temperature when these particles participate in the initiation of ice crystal formation and subsequently impact hailstone embryo growth. Biological particles in hailstones, such as pollen and bacteria, are more efficient ice-nucleating particles than dust within the ice nucleation region of storm clouds (Michaud et al., 2014). They can raise the freezing threshold temperature above −15 °C , while dust particles are activated to form ice crystals at temperatures below −15 °C (Michaud et al., 2014). In addition to species, number concentration of insoluble particles can also influence the hailstone formation. When more dust particles were considered , a model simulation resulted in larger number concentration of ice crystals, smaller graupels (one type of hailstone embryos) size, and suppression ~~in~~of the hailstone growth (Chen et al., 2019). Nonetheless, previous studies involving analysis of insoluble particles in hailstones mainly focused on substances analysis or total number concentration statistics. A size distribution of insoluble particles in hailstones with species information, which is beneficial for completing microphysical observation in severe storms, has not been given so far.

This study ~~identified~~analyzed insoluble particles ~~present~~ in hailstones~~, which were~~ collected from 8 hailstorms that occurred in China between 2016 and 2021~~, by scanning electron microscopy~~. The identification of insoluble particles in hailstones was conducted using Scanning Electron Microscopy (SEM) and ~~energy dispersive~~Energy Dispersive X-ray spectrometry (EDX). The~~se~~ insoluble particles were ~~grouped~~identified into three species ~~by self-organized maps (SOM~~using Self-Organized Maps (SOMs) and the random forest method. ~~Variation of~~The variation in size distribution of insoluble particles in embryos and different shells was explored. Based on ~~these analysis data~~the size distributions, logarithmic normal distributions were fitted to describe ~~different species~~the concentration of ~~insoluble particle~~organics and dust in deep convection.

Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted

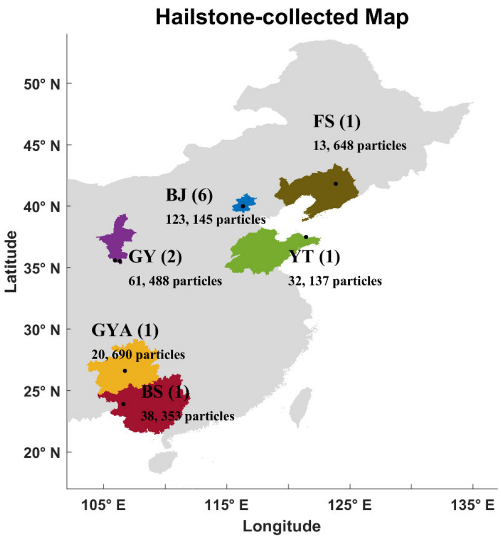
2 Methods

2.1 Sample information and experimental design

Formatted: Font: 小四

Hailstones were collected from eight hailstorms ~~occurring~~that occurred in six provinces of China during warm seasons from 2016 to 2021~~, and~~ (Table 1, Fig. 1). Volunteers stored the hailstones in clean containers, ~~such as~~including plastic bags, glass containers, and tinfoil, ~~by volunteers~~either during or ~~just~~immediately after the hail ~~(Table. 1, Fig. 1).~~events. All hailstone samples were transported to a laboratory at Peking University in Beijing and ~~stored~~kept at temperatures ~~between -~~ranging from -18 °C ~~and -~~to -4 °C. The hailstones were then transferred into vacuum-sealed plastic pockets and ~~kept~~preserved in a freezer, ~~with the~~ maintaining an internal temperature ~~maintained between -~~ranging from -29°C ~~and -~~to -23°C, until they underwent further processing and analysis.

Formatted: Font: 小五



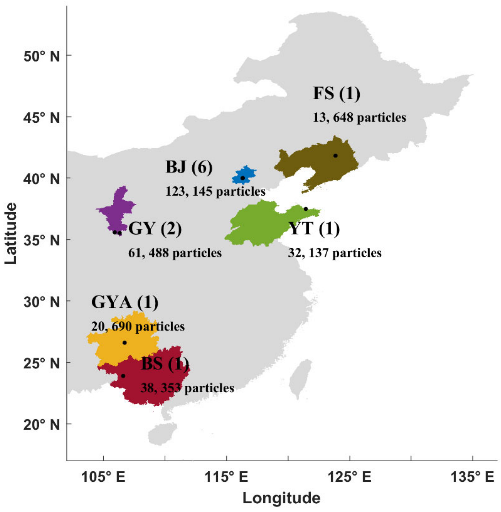


Fig. 1: Geographical distribution of ~~hailstone~~ collected ~~provinces. Black dots are~~ hailstones. The collecting locations of hailstones~~.~~ are indicated by black dots. Provinces of China from which the hailstones were collected are ~~shown in~~represented by six different colors. ~~The sample abbreviations are marked in the figure with~~The number of hailstones ~~sampled~~we analyzed was indicated in

parentheses. Abbreviations (corresponding to Table~~.~~ 1): BJ~~, Beijing City; GY, Guyuan City~~ - BeiJing; BS~~, Baise City~~ - BaiSe; FS~~, Fushun City; YT, Yantai City~~ - FuShun; GY - GuYuan; GYA~~, Guiyang City~~ - GuiYAng; YT - YanTai.

Table~~.~~ 1: Information about collected hailstones.

Date & Beijing Local ~~Solar~~ Time[a]	~~Location~~[b]Latitude & Longitude[c]	Total column water vapor[d] (kg / m^2)	Freezing level height – orography altitude[e] (m)	~~City~~Location & Sample abbreviation[f]	Samples[g]	Diameter[h] (mm)	Particle number[i]
19 June 2018, 18:30	41.82° N, 123.85° E	26.35~~9~~[18]	3241.66[18]	~~Fushun City~~FuShun (FS)	1	13.80	13,~~-~~648
10 June 2016, ~~14:30~~15:00	40.00° N, 116.32° E	36.86[14]	3780.52[14]	~~Beijing City~~BeiJing (BJ1)	1	—	~~123,~~ ~~145~~35,291
30 June 2021, ~~19:00~~20:18	39.95° N, 116.30° E	31.~~73~~[~~18~~84[20]	~~3854.52~~[~~18~~3852.76[20]	~~Beijing City~~BeiJing (BJ2)	5	25.38	14,865
				~~Beijing City~~BeiJing (BJ3)		24.11	20,233
				~~Beijing City~~BeiJing (BJ4)		16.30	20,350
				~~Beijing City~~BeiJing (BJ5)		14.86	14,350
				~~Beijing City~~BeiJing (BJ6)		22.80	18,056
01 Oct 2021, 14:02	37.49° N, 121.44° E	32.81[13]	3642.42[13]	~~Yantai City~~YanTai (YT)	1	45.00	32,~~-~~137
25 Aug 2020, ~~17~~18:00	35.53° N, 106.32° E	17.83[~~16~~17]	422.58[~~16~~17]	~~Guyuan City~~GuYuan (GY1)	1	15.00	~~61,~~ ~~488~~29,341
26 Aug ~~2022,~~ ~~15~~2020, 16:00	35.58° N, 105.93° E	17.01[~~14~~15]	835.04[~~14~~15]	~~Guyuan City~~GuYuan (GY2)	1	18.50	32,107
14 Apr 2016, ~~19~~20:00	26.60° N, 106.72° E	31.62[~~18~~19]	2147.58[~~18~~19]	~~Guiyang City~~GuiYAng (GYA)	1	26.20	20,~~-~~690
09 May 2016, ~~17~~18:51	23.90° N, 106.60° E	47.45[~~17~~18]	4572.70[~~17~~18]	~~Baise City~~BaiSe (BS)	1	—	38,~~-~~353

Formatted: Font color: Text 1
Formatted: Position: Vertical: 11.99 ch, Relative to: Page
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Position: Vertical: 11.99 ch, Relative to: Page
Formatted: Position: Vertical: 11.99 ch, Relative to: Page
Split Cells
Formatted: Position: Vertical: 11.99 ch, Relative to: Page
Formatted: Position: Vertical: 11.99 ch, Relative to: Page
Formatted: Position: Vertical: 11.99 ch, Relative to: Page
Formatted: Position: Vertical: 11.99 ch, Relative to: Page
Split Cells
Formatted: Position: Vertical: 11.99 ch, Relative to: Page
Formatted: Position: Vertical: 11.99 ch, Relative to: Page
Formatted: Position: Vertical: 11.99 ch, Relative to: Page

[a] Date and Beijing local solar time of hailstorms occurrences. Hailstones were collected within 30 min during hail.

[b] Hailstone collecting location.

[c] Latitude and longitude where the hailstone were collected.

[d] The total Total column water vapor values (Beijing local solar time of ERA5 reanalysis data in square brackets (Hersbach et al., 2018)).

[e] Depth between freezing level height and orography (Beijing local solar time of ERA5 reanalysis data in square brackets(Hersbach et al., 2018)).

[f] Sample Location and sample abbreviations.

[g] Numbers of hailstones used in the experiments.

[h] Diameter of hailstone (——— means no record).

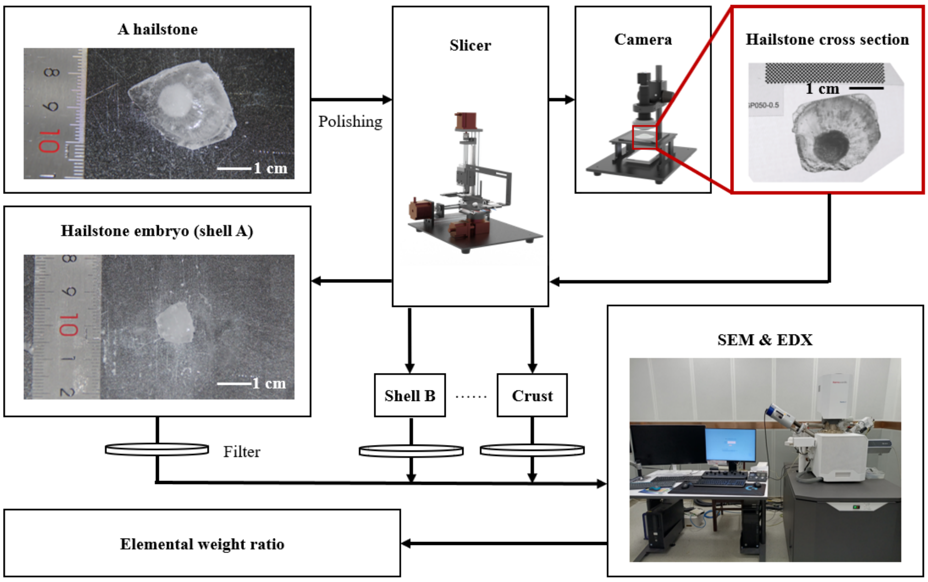
[i] Insoluble particle number in hailstones from the same province.

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Insoluble particles were extracted in the experiments (Fig. 2). The surface of ~~the~~each hailstone was polished to remove any attached grass or soil. ~~Then,~~Subsequently, the hailstones were sliced into cross-sections along the major axis, corresponding to the size of the hailstone embryo. The cross-section~~s~~ were further sliced into ~~several~~ shells using heated Fe-Cr alloy wire at an air temperature below −8°C. The shells within a hailstone were distinguished based on their natural transparency or opacity. ~~Hailstones~~However, hailstones with a major axis < 7 mm could not be sliced ~~because of~~due to the mass loss ~~with~~resulting from heating using our experimental apparatus.



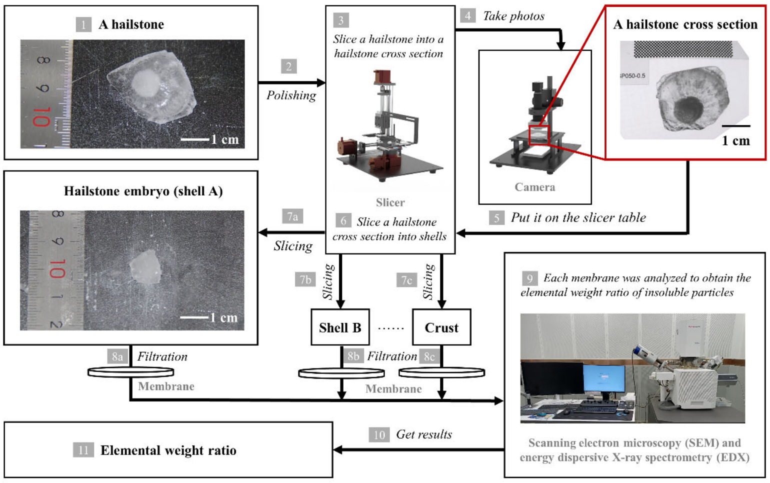


Fig. 2: Schematic diagram ~~showing~~illustrating the experimental framework. [1-2] The surface of ~~the~~each hailstone was polished to remove any attached grass or soil~~ and~~. [3] Subsequently, the hailstones were sliced into cross-sections along the major axis~~. The shells within a~~, corresponding to the size of the hailstone embryo. [4-7] After photographing the hailstone cross-sections, they were further subdivided into shells using heated Fe-Cr alloy wire at an air temperature below −8°C. The shells were distinguished based on their natural transparency or opacity. ~~Solution~~[8] The solution of melting shell samples ~~run~~was then passed through a filter membrane to ~~obtain~~isolate the insoluble particles. [9] Each shell sample ~~was analyzed within about 4 hours by~~underwent analysis using scanning electron microscopy and energy-dispersive X-ray spectrometry ~~for~~to determine the elemental weight ratios of the insoluble particles within approximately 4 hours. [11] Finally, the elemental weight ratio information of hailstones was obtained.

Formatted: No widow/orphan control

Formatted: Font: 小五, Bold

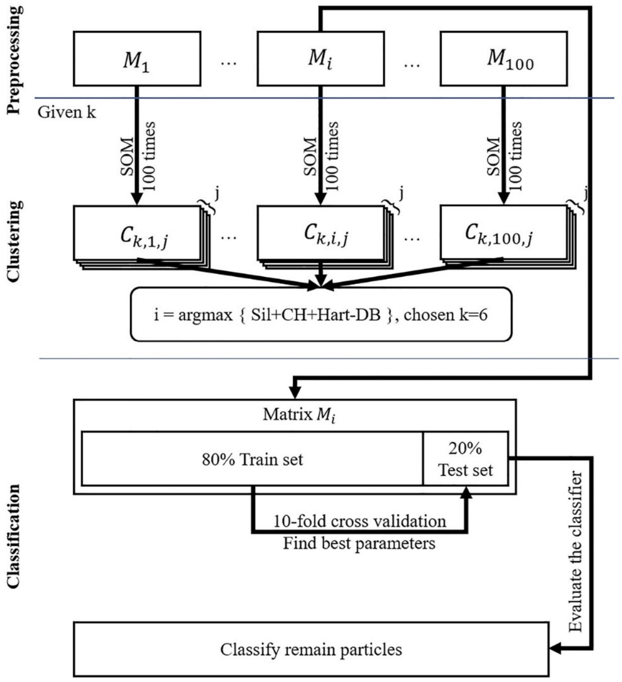
The shells were sequentially labeled with capital letters in alphabetical order, starting from the ~~inner shell to the crust. For example, the embryo of a hailstone was~~ embryo (designated as shell A~~. To obtain insoluble particles,~~) and progressing toward the crust. After the ice shells ~~were melted~~melting into a solution, ~~and run~~the solution was filtered through a ~~filter~~ membrane (VSWP01300, Merck KGaA, Germany) with a pore size of 30 nm. ~~The filter membrane was flushed five times with~~ The 1 mL (a total of 5 mL) of distilled water underwent five passes through the filter membrane to ensure ~~as many~~maximum retention of insoluble particles~~ as possible stuck~~ on the ~~filter~~ membrane. ~~The~~Subsequently, the filter membrane was dried under an air temperature of ~~about~~approximately 40°C ~~for electron microscopy~~to satisfy the dry-environment requirements of SEM.

The number of insoluble particles in each shell was determined ~~by~~using scanning electron microscopy (SEM), ~~focusing~~with a focus on particles ~~>~~ larger than 0.16 µm. The length along the major axis of the particles was measured using

Aztec software (Aztec software, Oxford Instruments plc, UK) on SEM images. ~~Energy~~ The software was able to randomly capture electron microscopy photos of the membrane (Aztec User Manual). No particle will be counted repeatedly. Energy-dispersive X-ray spectrometry (EDX) was ~~used~~utilized to determine the elemental weight ratios of the particles. Only elements with an atomic number ~~>~~ greater than 4 could be detected ~~becausee~~due to the X-ray input window ~~was~~being made of beryllium. Each shell sample was analyzed within ~~about~~approximately 4 hours by SEM and EDX. The scanning mode of SEM was set in a random order to reduce ~~the error~~errors caused by bias in the detection area.

2.2 Clustering and classification

The number of insoluble particles was measured using Aztec on SEM images, but the species could not be determined directly and were identified by machine learning method. The criteria of species classification were established by the ~~self-organized maps~~SOMs method to determine the species of unclassified particles. These labeled particles were then regarded as ~~true species and used to train a~~trainning set in random forest classifier. Details are presented in Fig. 3.



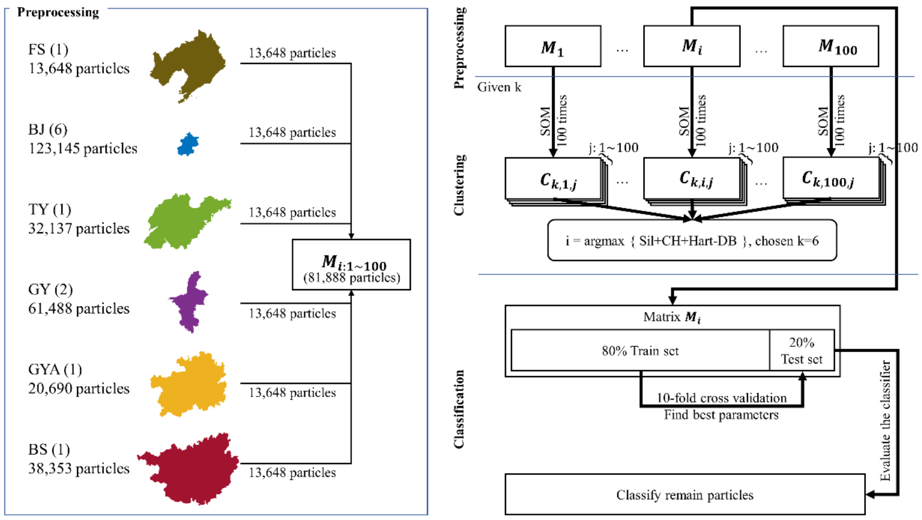


Fig. 3: Schematic diagram ~~showing~~illustrating the methodological framework ~~of the methods~~ used for particle identification in this study. ~~The~~A total of 100 matrices M_i, with i ranging from 1 to 100, were ~~used~~utilized in self-organized maps clustering analyses, ~~and~~ each ~~of them included~~ containing 81,888 unidentified ~~81,888~~ particles with 19 elemental features (N, Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, Cr, Mn, Fe, Ni, Cu, Br, Ba, and Pb). ~~Centroid~~The centroid matrix $C_{k,i,j}$ ~~was~~ represents the clustering results ~~by~~ obtained through the self-organized maps method with ~~chosen~~a given cluster number k. The ~~operation of~~ self-organized maps operation with the same k was repeated 100 times to ensure ~~the~~result robustness ~~of results. The~~, where j ~~is~~denotes the number of ~~repeating time,~~repetitions ranging from 1 to 100. Four ~~indexes, i.e.,~~indices, Silhouette index (Sil), Calinski–Harabasz index (CH), modified Hartigan index (Hart), and Davies–Bouldin index (DB~~)~~), were ~~used~~employed to determine ~~best centroid number~~the optimal parameters k, i, and j. The matrix M_i containing identified 81,888 particles~~,~~ was ~~separated as~~randomly divided into a training set (80 %) and a test set ~~in~~(20 %) for random forest classification ~~with~~. The 10-fold cross-validation~~. The best classifier~~ was ~~used~~utilized to ~~classify remain particles~~determine the best tree. Abbreviations (corresponding to Table 1): BJ - BeiJing; BS - BaiSe; FS - FuShun; GY - GuYuan; GYA - GuiYAng; YT - YanTai.

Formatted: Heading 1, Justified, Line spacing: single

~~With reference to the studies of Ault et al. in 2012 and Kirpes et al. in 2018 and considering the results of elemental weight ratios determined by EDX analysis (Ault et al., 2012; Kirpes et al., 2018), 19 elements (N, Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, Cr, Mn, Fe, Ni, Cu, Br, Ba, and Pb) were selected to confirm the species of particles. C and O were not taken in account when clustering or classifying particles as the membrane filters were made from cellulose acetate and cellulose nitrate, which contain C, H, N, and O. We could not detect H because the ray input window was made of beryllium. All particles showed high contents of C and O but different contents of N, so N was retained as a feature of classification.~~

With reference to the studies of Ault et al. (2012) and Kirpes et al. (2018) and considering the results of elemental weight ratios determined by EDX analysis, 19 elements (N, Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, Cr, Mn, Fe, Ni, Cu, Br, Ba, and Pb) were selected to confirm the species of particles. C and O were not taken in account when clustering or classifying particles as the membrane filters were made from cellulose acetate and cellulose nitrate, which contain C, H, N, and O. We could not detect H because the ray-input window was made of beryllium. All particles showed high contents of C and O but different contents of N, so N was retained as a feature of classification.

Species of aerosol particles vary ~~regionally~~with sampling location (Tao et al., 2017). Therefore, when establishing the matrices of elemental weight ratios for clustering, equal amounts of data were randomly extracted from the sample data from each province to ensure the inclusion of a consistent proportion of samples from each region in the training process. A hailstone FS collected from Fushun City, Liaoning Province was shown to contain 13,648 insoluble particles, which was the smallest among all samples from six provinces (Fig. 1). With random sampling of 13,648 particles from each province, the matrix used in clustering analyses included 81,888 particles. This operation was repeated 100 times to obtain 100 matrices $\mathbf{M_i}$ with i ranging from 1 to 100.

Each matrix $\mathbf{M_i}$ was clustered using the ~~SOM method, which is an unsupervised machine learning method that represents high-dimensional data in low-dimensional space while preserving the topological structure of the data. The neuronal network was set to k neurons in a layer, where k is the given clustering center number from 2 to 10. Each SOM operation produces a centroid matrix $\mathbf{C_{k,i,j}}$, where i is the number of particle sample replicates, as mentioned above, and j is the number of rounds of SOM operation. Weights of a neuron describe its position in multivariate space and can be taken as a cluster center. The operation of SOM with the same neuronal network setting was repeated 100 times to ensure the robustness of the centroid matrix $\mathbf{C_{k,i,j}}$. Four indexes, i.e., Silhouette index~~SOMs method. SOMs belong to the category of competitive learning algorithms and are a type of artificial neural network (Kohonen, 1990). A basic SOMs network consists of an input layer, weight vectors, and an output layer. Each neuron in the output layer possesses a set of weight vectors, which represent the topological structure of the neurons in the output layer, associated with the inputs. SOMs are commonly used as dimensionality reduction algorithms, enabling the representation of high-dimensional data in a lower-dimensional structure while preserving the original topology. When SOMs are trained on unlabeled data for clustering purpose, it proves highly beneficial in clustering unlabeled and high-dimensional inputs into visualized two-dimensional outputs.

We utilized the SOMs code from MATLAB's deep learning toolbox. The input of SOMs is $\mathbf{M_i}$. At begin, the neural network in the output layer was initialized as 1-D dimension with k neurons. The number of neurons in the output layer matches k ranging from 2 to 10. The operation of SOMs with the same initialized k neurons and input matrix $\mathbf{M_i}$ was repeated 100 times to ensure result robustness. The clustering result was stored in matrix $\mathbf{C_{k,i,j}}$, which corresponded to the

Formatted: Font: 五号

Formatted: Don't add space between paragraphs of the same style

Formatted: Don't add space between paragraphs of the same style given k centroids in $\mathbf{M_i}$ with j^{th} SOMs operation. Each $\mathbf{C_{k,i,j}}$ matrix consists of k rows and 19 columns (corresponding to the number of elemental features). Four indices, namely, the Silhouette index (Rousseeuw, 1987), the Calinski–Harabasz index (Calinski and Harabasz, 1974), the modified Hartigan index (Sibson and Hartigan, 1976), and the Davies–Bouldin index (Davies and Bouldin, 1979), were selected as evaluation indicators to determine the parameters k, i and j. ~~The best k, i and j was chosen by combining the evaluation of the four indexes (Fig. 4) and elemental weight ratios of each centroid~~The Silhouette index, Davies–Bouldin index, and Calinski–Harabasz index assess the similarity between a particle and others within the same cluster, as well as the dissimilarity across different clusters for a given k. Hartigan index evaluates whether it is worthy to increase the k. Notably, Hartigan index has undergone modifications that preserve its statistical meaning while conserving computational resources.

Hartigan index (Sibson and Hartigan, 1976) is defined as:

$$H(k) = (N - k - 1)\left[\frac{err(k)}{err(k+1)} - 1\right], k = 2\sim10 \quad (1)$$

$$err(k) = \sum_{g=1}^{k}\sum_{x_g \in C_g}\left(x_g - C_g\right)^2 \quad (2)$$

k : the number of clusters.

C : the centroid of all data

N : the number of observations in data

C_g : the centroid of cluster g

x_g : the observation of cluster g

x_n : the observation of data

The calculation of $H(k)$ requires clustering for values of k ranging from 2 to 11 in order to obtain $H(2)$, $H(3)$, …, $H(10)$. Clustering particles into 11 clusters would require performing an additional 10,000 iterations of the SOMs, with 100 iterations of extracting $\mathbf{M_i}$ and 100 iterations of SOMs for each $\mathbf{M_i}$. Additionally, we observed that the SOMs did not perform well in the Silhouette index (Sil), the Calinski–Harabasz index (CH), and the Davies–Bouldin index (DB) when $k = 2$. As a result, we introduced modifications to the Hartigan index.

$$Hart(k) = [N - (k - 1) - 1]\left[\frac{err(k-1)}{err(k)} - 1\right], k = 2\sim10 \quad (3)$$

$$err(k) = \sum_{g=1}^{k}\sum_{x \in C_g}\left(x_g - C_g\right)^2, k \geq 2 \quad (4)$$

When $k = 1$, it indicates that all particles are belong to one cluster.

$$err(1) = \sum_{n=1}^{N} (x_n - C)^2 \tag{5}$$

In clustering with a specific value of k, our objective is to have particles tightly grouped together in feature space while ensuring that the centroids exhibit a significant dispersion compared to $k-1$. A higher value of $Hart(k)$ for a given k indicates improved clustering performance. The best k, i and j was chosen by combining the evaluation of the four indices (Fig. 4). We applied max normalization to rescale the four indices, $Sil(k)$, $CH(k)$, $DB(k)$, and $Hart(k)$. Subsequently, the best combination of k, i and j was determined, resulting in $\{Sil(k,i,j) + CH(k,i,j) + Hart(k,i,j) - DB(k,i,j)\}$ reaching its maximum.

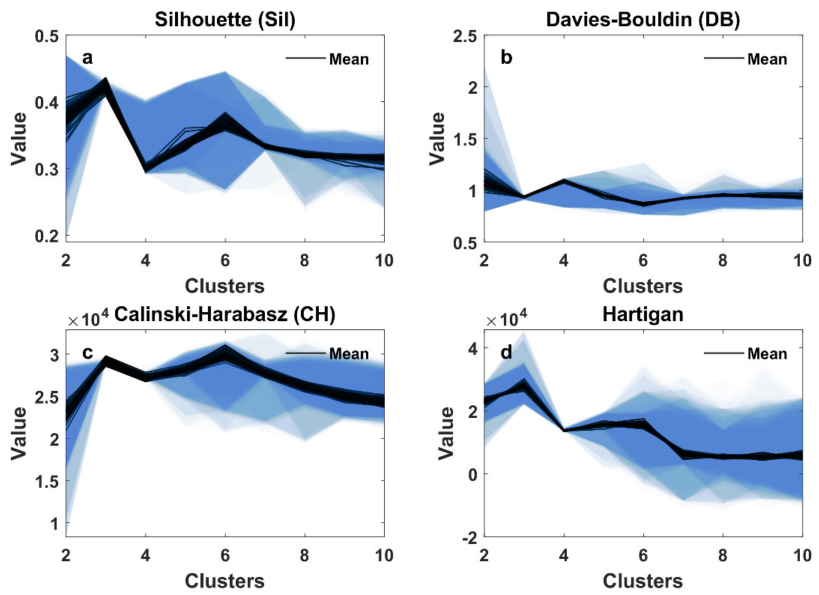
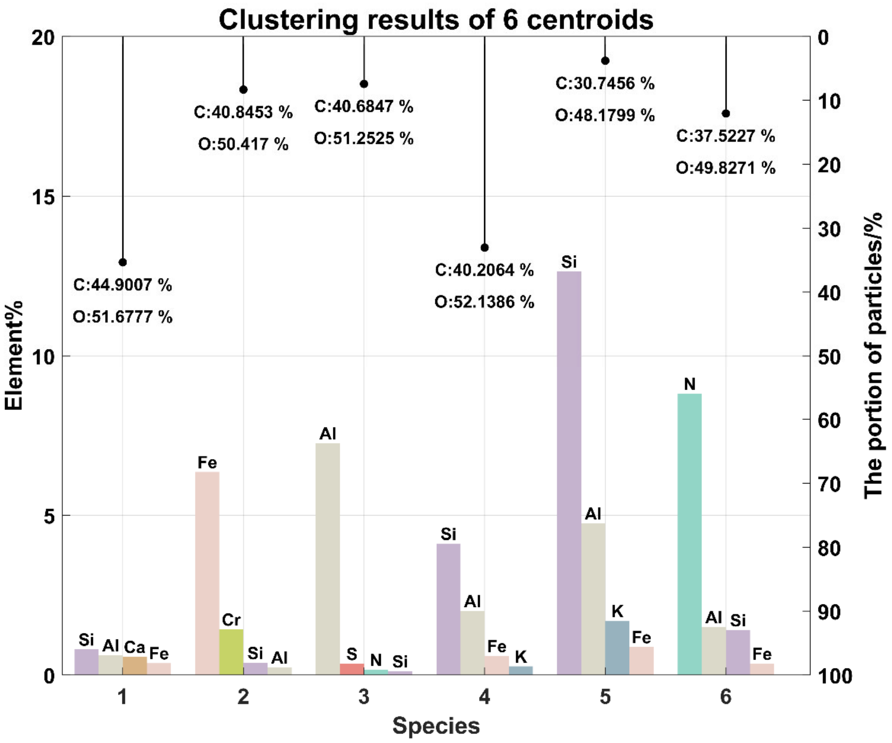


Fig. 4: Evaluation of self-organized maps clustering results. ~~Evaluation~~The clustering results of self-organized maps ~~clustering results by~~were evaluated using (a) Silhouette index, (b) Davies–Bouldin index, (c) Calinski–Harabasz index, and (d) Hartigan index. ~~Self~~The self-organized maps operation was repeated 100 times to ~~obtain each randomly sampled matrix M_i.~~ensure result robustness. The solid lines and shading represent the average and spread of 100 repetitions, respectively.

Formatted: Indent: First line: 0.74 cm

The centroid matrix ~~$C_{k,i,j}$~~$\mathbf{C_{k,i,j}}$ with best k, i and j was treated as a training set for random forest classification. The chosen centroid matrix ~~$C_{k,i,j}$~~$\mathbf{C_{k,i,j}}$ with the top four elements is shown in Fig. 5 with $k = 6$. The first species with low elemental weight ratio except C and O contents was considered to be organics. The second species with high Fe content and low Cr content was introduced by the material of the slicer used in the experiment. The third species had a high Al content representing oxides or carbonates of aluminum. The fourth and fifth species were mineral silicates. So that, the third, fourth, and fifth species were referred to as “dust”. The last species with high N content was protein-containing biological aerosol.



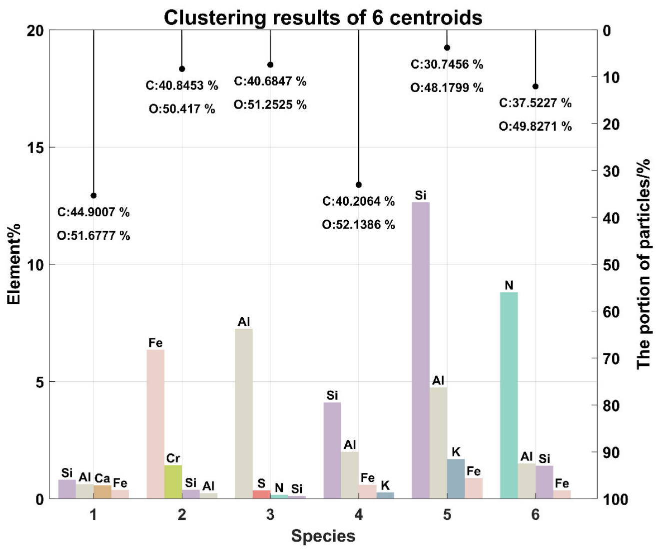
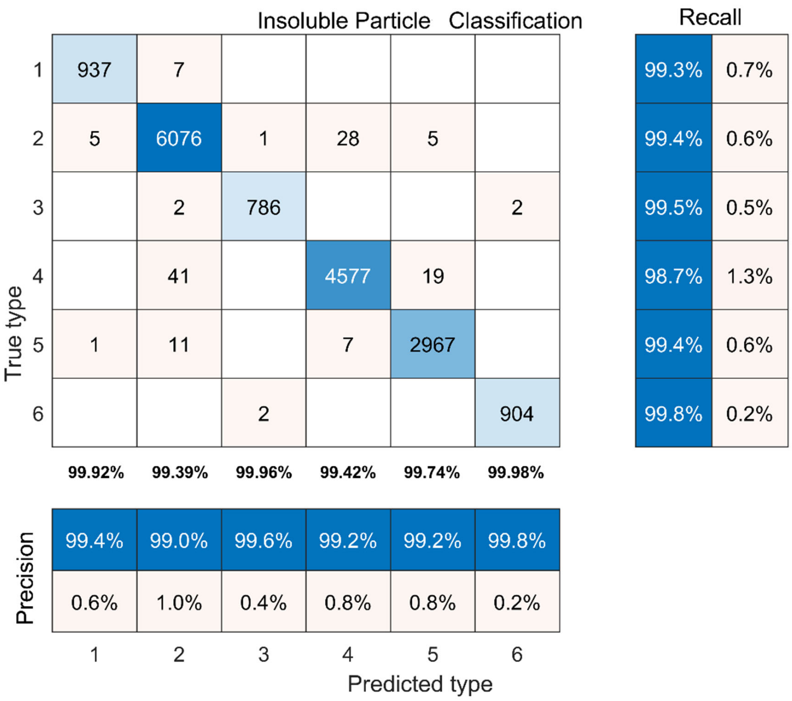


Fig. 5: Centroids of clustering with six clusters from self-organized maps results and each species portion. Colored bars show the top four elements of each species. The stem bars show the portion of each species. The average contents of C and O of each species are marked at the end of the stem bars.

The random forest method was applied in classifying insoluble particles, which involves randomly growing 100 classification trees. The training set consisted of 80 % of $\mathbf{M_i}$ and 10-fold stratified cross-validation was applied during the training process to find the best tree among the 100 random trees. The remaining 20~~%~~ % particles of $\mathbf{M_i}$ was used as the test set to evaluate the ~~classifier. The~~ best ~~classification~~ tree~~ and the~~. The confusion matrix of ~~the evaluation of testing~~classification results are shown in Fig. 6. All remaining insoluble particles were classified by this tree. Finally, ~~results~~we identified three species: organics, dust, and bioprotein aerosols.



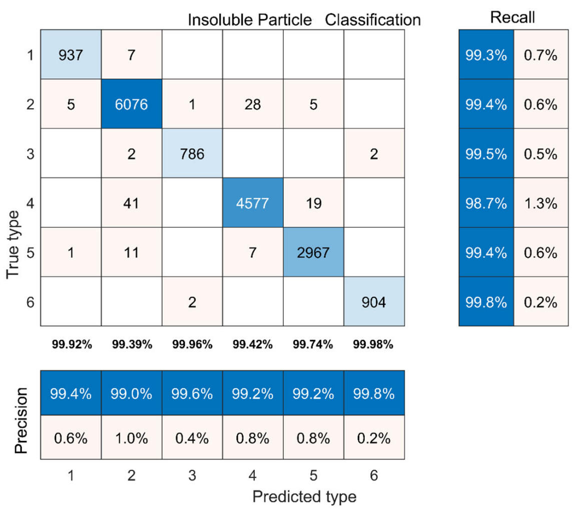


Fig. 6: Confusion matrix of the best random forest classifier tree. The numbers on the diagonal are accurately predicted insoluble particles. Numbers in bold indicate the accuracy of prediction of each type.

2.3 ~~Calculation~~Conversion of insoluble particle~~s~~ number concentration

Particle number was converted to a number concentration ~~as follows~~per cubic centimeter volume water (hereinafter referred to as number concentration) using the following formula:

~~$$n_{liquid} \cdot V_{liquid} = N_{liquid} = N_{dilute} = n_{diluted} \cdot V_{diluted} \quad (1)$$~~

~~Part of the solution was not consumed in these experiments and was retained as a backup. During several experiments, the melted shell solution was diluted.~~

~~$$n_{diluted} = n_{used} = \frac{N_{used}}{V_{used}} = \frac{N_{filter}}{V_{used}} \quad (2)$$~~

~~*SEM can provide the number of particles on a filter, but the whole area of the filter cannot be scanned. We assumed*~~

$$n_{liquid} \cdot V_{liquid} = N_{liquid} = N_{diluted} = n_{diluted} \cdot V_{diluted} \quad (6)$$

The number of insoluble particles in the melted shell solution (N_{liquid}) can be calculated by multiplying their number concentration (n_{liquid}) with the volume of the shell solution (V_{liquid}). Part of the solution was not used up in the experiments and was kept as a backup. Therefore, the shell solution was diluted in some experiments and part of the solution was consumed in the experiments. As in the melting solution, the number of insoluble particles in the diluted solution ($N_{diluted}$) can be calculated by multiplying their number concentration ($n_{diluted}$) with the volume of the diluted solution($V_{diluted}$). The total particle number in the melted shell (N_{liquid}) remains unchanged during the dilution process ($N_{diluted}$).

$$n_{diluted} = n_{used} = \frac{N_{used}}{V_{used}} \quad (7)$$

The number concentration of the diluted solution ($n_{diluted}$) is equal to that ~~the~~ of the consuming part (n_{used}). Assuming the rinsing operation ensures all insoluble particles in the shell were ~~uniformly distributed~~ on the ~~filter and the scanning mode of SEM was set as "random scanning". A such relationship between~~membrane, the number of ~~scanned~~insoluble particles ~~and the number of particles on the filter:~~

~~$$\frac{S_{filter}}{S_{images}} = \frac{N_{filter}}{N_{count}} \quad (3)$$~~

~~In the above formulas, n is the number concentration of insoluble particles; N~~in the consumed solution (N_{used}) is ~~the number of insoluble particles; V is the volume of the solution; S is the area of the filter; subscript *liquid* refers to the melted shell; subscript *diluted* refers to the diluted solution; subscript *used* refers~~equal to the ~~consumed diluting solution; subscript *filter* refers to the filter membrane; N_{count} is the number of~~ number of insoluble particles counted on the ~~filter; and S_{images} is the area of the microscopic image.~~membrane (N_{filter}).

We use SEM to capture electron microscopy images of the membrane. Assuming a uniform distribution of insoluble particles on the filter membrane, a software randomly capture electron microscopy photos of the membrane and count the visible insoluble particles in those images. The relationship between total number of visible insoluble particles counted in the images (N_{count}) and N_{filter} is:

$$\frac{S_{images}}{S_{filter}} = \frac{N_{count}}{N_{filter}} \tag{8}$$

That is, N_{filter} is determined by multiplying N_{count} by the ratio of the areas between the entire filter membrane (S_{filter}) and the electron microscopy images (S_{images}). These three formulas Eq. (6-8) were reduced to Eq. (~~4~~9):

~~$$n_{liquid} = \frac{1}{V_{liquid}} \cdot \frac{S_{filter}}{S_{images}} \cdot \frac{V_{diluted}}{V_{used}} \cdot N_{count} \tag{4}$$~~

~~*where*~~

$$n_{liquid} = \frac{1}{V_{liquid}} \cdot \frac{S_{filter}}{S_{images}} \cdot \frac{V_{diluted}}{V_{used}} \cdot N_{count} \tag{9}$$

Here, $S_{filter}, S_{images}, N_{count}, V_{diluted}$, and V_{used} can be measured. The liquid volume (V_{liquid}) was determined as the ~~mean~~average of readings obtained by two experimenters from the test tube~~ calibration. From Eq. (4), a tiny change in n_{liquid} can be expressed as dn_{liquid}:~~. Take the logarithm on both sides:

~~$$dn_{liquid} = n_{liquid} \cdot \left(-\frac{dV_{liquid}}{V_{liquid}} + \frac{dV_{diluted}}{V_{diluted}} - \frac{dV_{used}}{V_{used}} + \frac{dN_{count}}{N_{count}} \right) \tag{5}$$~~

$$\ln n_{liquid} = -\ln V_{liquid} + \ln S_{filter} - \ln S_{images} + \ln V_{diluted} - \ln V_{used} + \ln N_{count} \tag{10}$$

Based on Eq. (10), a tiny change in n_{liquid} can be represented as dn_{liquid}:

$$dn_{liquid} = n_{liquid} \cdot \left(-\frac{dV_{liquid}}{V_{liquid}} + \frac{dV_{diluted}}{V_{diluted}} - \frac{dV_{used}}{V_{used}} + \frac{dN_{count}}{N_{count}} \right) \tag{11}$$

As,

~~$$dS_{filter} = dS_{images} = 0 \tag{6}$$~~

~~The uncertainty comes from the measurement error of the experimental instruments.~~

~~$$\Delta = |dn_{liquid}| \leq n_{liquid} \cdot \sqrt{\left(\frac{dV_{liquid}}{V_{liquid}}\right)^2 + \left(\frac{dV_{diluted}}{V_{diluted}}\right)^2 + \left(\frac{dV_{used}}{V_{used}}\right)^2 + \left(\frac{dN_{count}}{N_{count}}\right)^2} \tag{7}$$~~

~~So,~~

~~$$\Delta_{max} = n_{liquid} \cdot \sqrt{\left(\frac{dV_{liquid}}{V_{liquid}}\right)^2 + \left(\frac{dV_{diluted}}{V_{diluted}}\right)^2 + \left(\frac{dV_{used}}{V_{used}}\right)^2 + \left(\frac{dPs}{Ps}\right)^2} \tag{8}$$~~

$$dS_{filter} = dS_{images} = 0 \tag{12}$$

The uncertainty (Δ) of n_{liquid} comes from the measurement error of the experimental instruments, following below (Taylor, 1997):

$$\Delta = n_{liquid} \cdot \sqrt{\left(\frac{dV_{liquid}}{V_{liquid}}\right)^2 + \left(\frac{dV_{diluted}}{V_{diluted}}\right)^2 + \left(\frac{dV_{used}}{V_{used}}\right)^2 + \left(\frac{dN_{count}}{N_{count}}\right)^2} \tag{13}$$

Here, the ~~minimum scale~~accuracy of the test tube ~~containing melting solution~~ is 0.1 mL~~ and~~. The term dV ~~is~~represents the greatest reading error caused by human and was set to 0.05 mL. The quantity $\frac{dN_{count}}{N_{count}}$ ~~represents~~corresponds to the uncertainty ~~of detecting~~associated with size of insoluble particles~~, which is related to~~ and the scan settings.

~~$$\frac{dN_{count}}{N_{count}} = \frac{dPs}{Ps} = \frac{3}{6{,}340{,}608} \tag{9}$$~~

$$\frac{dN_{count}}{N_{count}} = \frac{dPs}{Ps} = \frac{3}{6{,}340{,}608} \quad (14)$$

The term dPs ~~is~~represents the minimum number of pixels that can be detected in an image. Ps ~~is~~denotes the total number of pixels in the micrograph.

2.4 Curves fitting

We aggregated ~~our data~~insoluble particles into 0.2-μm intervals (~~e.g., particle number concentration at D = 0.3 μm, corresponding to the sum of particles of diameter 0.2–0.4 μm~~0.2 μm bin interval in Fig. 7 and Fig. 10, and 2 μm bin interval in Fig. 8 and Fig. 9) to fit the logarithmic normal distribution:

~~$$n(\ln D) = \frac{N}{\sqrt{2\pi}\ln\sigma_g}\cdot\exp\left[-\frac{(\ln D - \ln r_g)^2}{2\ln^2\sigma_g}\right] \quad (10)$$~~

~~$$n(D) = \frac{1}{D}\cdot n(\ln D) \quad (11)$$~~

~~Here,~~

$$n(\ln D) = \frac{N}{\sqrt{2\pi}\,ln\,\sigma_g}\cdot exp\left[-\frac{(ln\,D - ln\,D_g)^2}{2\,ln^2\,\sigma_g}\right] \quad (15)$$

N denotes the total number concentration of particles. Both $n(\ln D)$ and $n(D)$ ~~are~~represent the size distributions of particles, where D is the diameter of insoluble particles~~, and N is the total number concentration of particles. According to the above, when~~. $n(\ln D)$ and $n(D)$ can be converted to each other by D.

$$n(D) = \frac{1}{D}\cdot n(ln\,D) \quad (16)$$

When the N_{count} in an interval equals 1, the number concentration will ~~show~~exhibit a flat tail ~~because of~~due to the conversion to obtain n_{liquid}. The fitting data were selected with intervals equals to ~~0.2 μm~~0.2 μm. The least squares method was applied to determine the fitting parameters and R^2 was used to estimate ~~fitting parameters.~~the goodness of fit. The two centroids of fitting parameters of organics and dust were determined by K-means method.

3 Results

~~Total~~A total of 289,461 insoluble particles were detected from 30 shells of 12 hailstones ~~were detected by scanning electron microscopy. Elemental weight ratios of each particle were determined~~ using ~~energy dispersive X-ray spectrometry. More details regarding calculating number concentration~~SEM. The identification of insoluble particles ~~per cubic centimeter volume water (hereinafter referred to as number concentration) from number of insoluble particles were showed in method description. Identification of insoluble particles used self-organized maps~~employed SOMs for clustering and random forest for

Formatted: Font: Times New Roman, 10 pt

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font color: Auto

Formatted: Font: Not Italic

classification. Four indicexes were selectedutilized to determine the appropriate parameters ofin clustering. The clustering results ($C_{k,i,j}$) were set asdivided into a training and a testing set offor classification. AThe confusion matrix of the best classifier showed that thean accuracy, precision, and recall wereof 99.7 %, 99.4 %, and 99.5 %, respectively. All particles were identifiedclassified as organics, dust, and bioprotein aerosols (i.e., the fraction of biological aerosols with protein content).

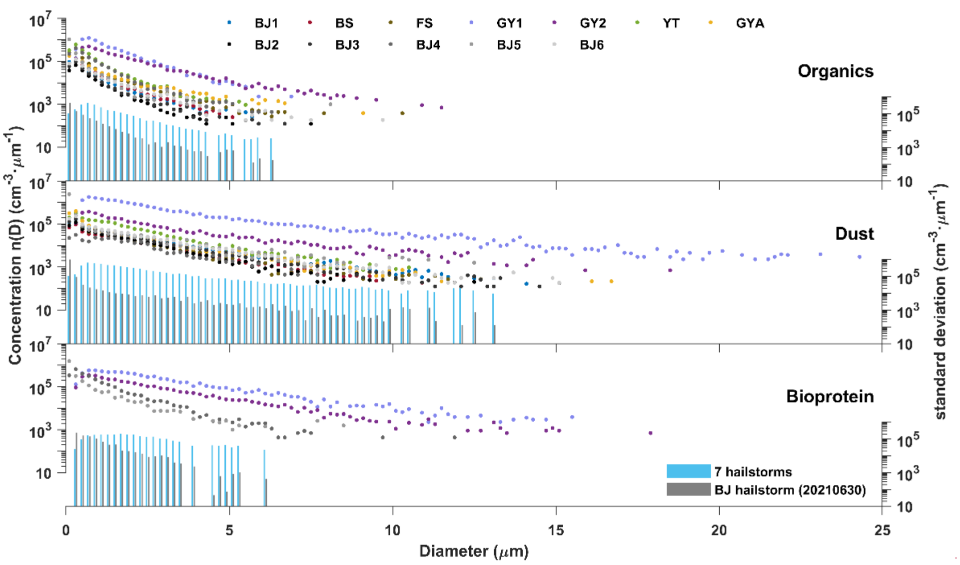
3.1 Sample representativenesssimilarity

Five of the 12 hailstones (BJ2–BJ6) were from the same hailstorm that occurred in Beijing on June 30, 2021. The insoluble particles present in these hailstones BJ2–BJ6 showed similarity in the size distribution of organics, dust, and bioprotein aerosols but differed, while those from other 78 hailstones that from other hailstorms(BJ1, BJ2, BS, FS,GY1, GY2, YT and GYA) exhibited a wider dispersion (Fig. 7). The results were similar to those of Li et al., who reported that the number concentrations of water-soluble ions varied among hailstorm events but showed similarity in the same storm (Li et al., 2018). These analyses suggested that insoluble particles in the hailstorm may come from local natural or anthropogenic emissions (e.g., soil dust, aerosols from biomass and fossil fuel combustion, products of the conversion of gaseous precursors), which is also suggested by the results on water-soluble ions (Beal et al., 2022). The updraft within the hailstorm is likely to bring insoluble particles from local surfaces or boundary layers into deep convective clouds, as hailstorms are among the most severe storms with strong updrafts (Battaglia et al., 2022). BJ2 was selected to represent five hailstones from the same hailstorm in further analysis to simplify comparison.

Formatted

Formatted

Formatted



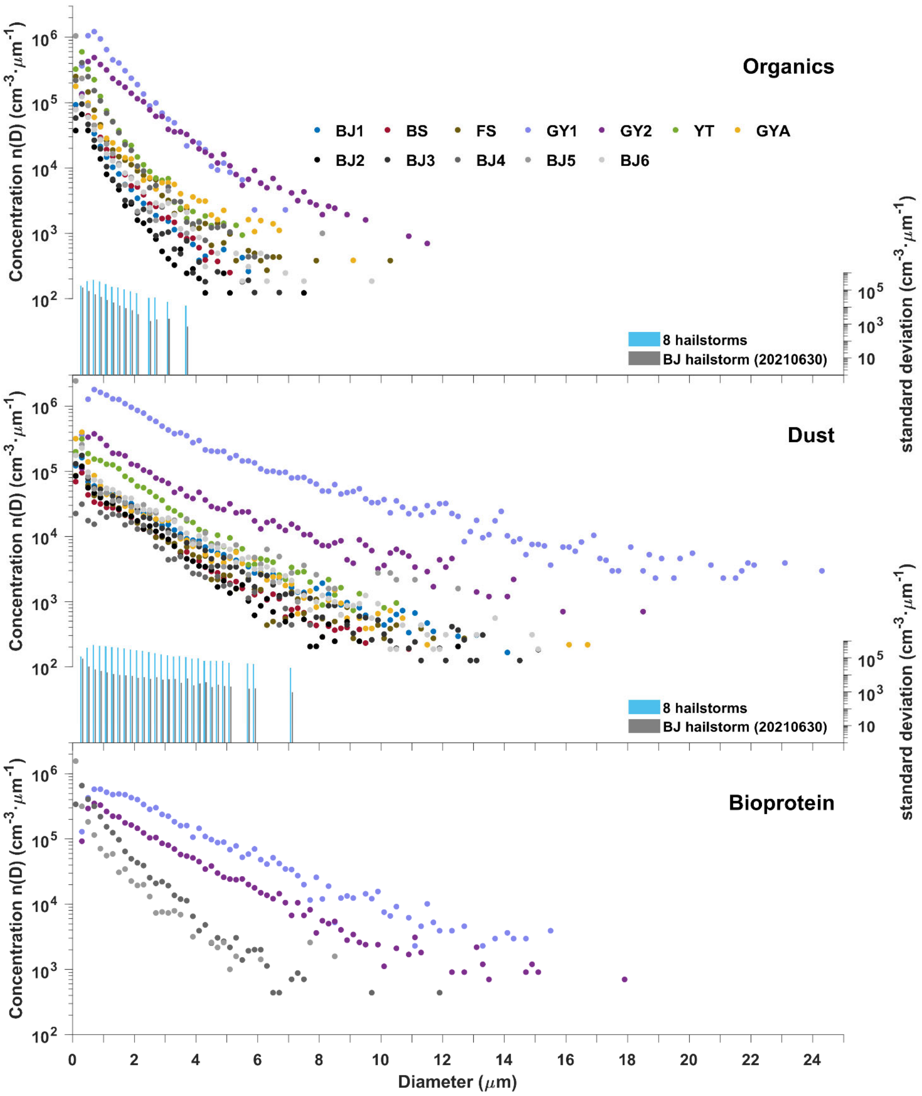


Fig. 7: Size distribution of organics, dust, and bioprotein aerosols of insoluble particles in 12 hailstones. ~~Each number concentration at diameter D total number concentration of insoluble particles with diameter ranging~~The colored dots represent data from ~~D – 0.1 μm to D + 0.1 μm. Colored dots refer to seven~~7 hailstones ~~(~~BJ1, BS, FS, ~~BJ-1~~GY1, GY2, YT~~, NX-1, NX-2, GY,~~ and ~~BS)~~GYA which were from ~~seven~~ different hailstorms. ~~Black~~The black and gray dots ~~refer~~correspond to ~~five~~data from hailstones (~~BJ-2–BJ-6)~~BJ2 to BJ6) that were from the same hailstorm ~~that occurred~~occurring in Beijing on June 30, 2021. ~~Blue~~The blue and gray bars

Formatted: Heading 1, Justified, Line spacing: single

~~show~~indicate the standard deviation of number concentration of insoluble particles from ~~seven hailstorms and one hailstorm~~8 hailstones (BJ1, BJ2, BS, FS,GY1, GY2, YT and GYA) from 8 cases and 5 hailstones (BJ2 to BJ6) from one case, respectively. Abbreviations (corresponding to Table~~.~~ 1): BJ~~, Beijing City; GY, Guyuan City~~ - BeiJing; BS~~, Baise City~~ - BaiSe; FS~~, Fushun City; YT, Yantai City~~ - FuShun; GY - GuYuan; GYA~~, Guiyang City~~ - GuiYAng; YT - YanTai.

3.2 Size distribution in embryos

Formatted: Space Before: 0 pt, After: 0 pt, Line spacing: 1.5 lines

All hailstone embryos analyzed in this study ~~were graupels~~are graupel particles, which grows from the initial ice particles through accretion of supercooled droplets (Knight, 1981). These initial ice particles are formed through nucleation of insoluble particles where heterogeneous nucleation take place (Lamb and Verlinde, 2011). In other words, insoluble particles in graupels influence the formation of ice crystals and subsequently affect the formation of hailstone embryos.

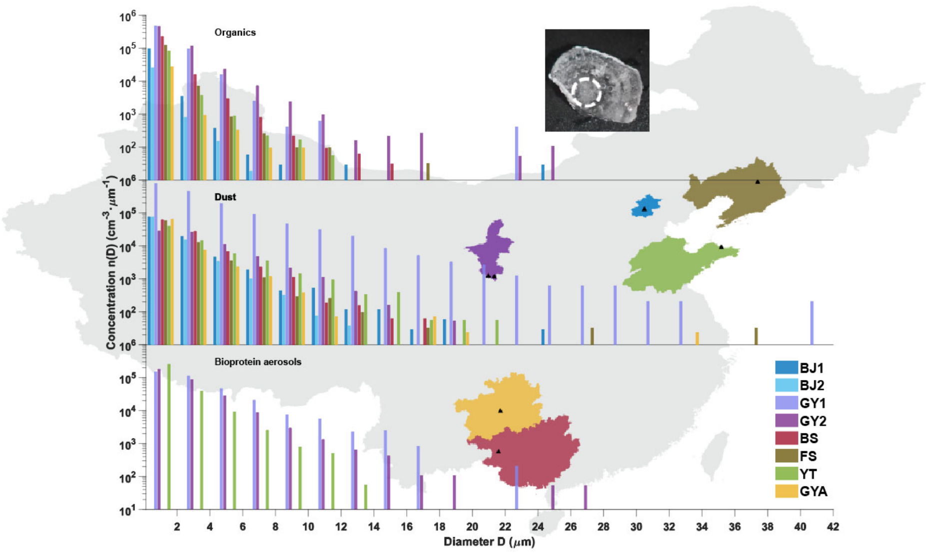


Fig. 8: Size distribution of insoluble particles in hailstone embryos. Different colors represent the provinces in China from which the hailstones were collected. Black triangles indicate the locations of hailstone sample collection. The white dashed circle highlights part of the hailstone embryo. Abbreviations (corresponding to Table 1): BJ - BeiJing; BS - BaiSe; FS - FuShun; GY - GuYuan; GYA - GuiYAng; YT - YanTai.

The variations in number concentrations of dust and bioprotein insoluble particles indicate that particle number concentrations decrease exponentially with particle diameter, with markble variation observed among hailstorms (Fig. 8). BJ2 was selected to represent five hailstones from the same storm to simplify comparison. The size distribution distinguishes organics from dust and bioprotein aerosols. The number concentrations of organics from all samples decrease with particle diameter less than 8 μm, while those of GY1 and GY2 fluctuate starting at diameters of 8 μm and 12 μm, respectively. Compared to other hailstones, GY1 and GY2 were collected in remote areas, where is fields of rural areas dedicated to growing crops near the south of the Gobi Desert. Therefore, GY1 and GY2 have a coarse mode of organics with particle diameters larger than 12 μm, possibly might due to the emission of spring-wheat straw burning and unrestricted diesel engine vehicles. The transport of coal combustion in surrounding cities may also contribute to the coarse mode organics. Among all cases, there is a significant variance in the size distribution of both organics and dust. The number concentration of organics from a hailstone embryo varied from 1 to 390 times, compared to those at the same particle diameter in hailstone embryos from different cases. The number concentration of dust from a hailstone embryo varied from 1 to 527 times, compared to those at the same particle diameter in hailstone embryos from different cases. The number concentrations of dust from BJ1, BJ2, and GY1 are at least 3 times higher than organics in particles of the same diameter in the range of 2–24 μm.

Moreover, dust showed a wider size distribution than organics and bioproteins among all samples. Dust from GY1 had a higher number concentration and larger maximum size (42 μm) compared to other hailstone embryos. Hailstone samples with high insoluble particle content, i.e., GY1 and GY2, showed significantly lower total column water vapor values and smaller depth between freezing level height and orography within one hour before hailstorm occurrence, compared to other hailstones (Table 1). The competition of condensation and relative shorter updraft pathway might be responsible for the high number concentrations of organics, dust, and bioproteins in GY1 and GY2, as compared with other haistones. Bioprotein aerosols, with high freezing efficiency, may have formed initial ice particles in GY1, GY2, and YT, while dust or organics formed initial ice particle in hailstorms in the other five cases. All hailstone embryos contained organics and dust, but not all hailstone embryos contained a significant amount of bioprotein aerosols. Due to limited comprehension of the transportation and transformation processes of biological materials, it is challenging that to establish a definitive relationship between biological protein particles and biological aerosols (Fröhlich-Nowoisky et al., 2016).

3.3 Size distribution in shells

Size distribution of each species varieded little in characteristics between outer shells with the embryos (Fig. 9). In a four-shell hailstone, the number concentrations of insoluble particles exhibited V-shaped distributions (BS and YT) or inverse V-shaped distributions (BJ1) from embryo to crust. Five of nine two-shell hailstones showed higher number concentrations of dust in crusts than embryos, while seven of them showed higher number concentrations of organics in embryos than crusts. These initial ice particles are likely formed by insoluble particles where heterogeneous nucleation processesMoreover, the quantification of differences in number concentration varied little among shells. The 90.5 % points showed that differences in number concentration of the same kind particles in a shell compared to the previous shell at the same diameter was within twice (294 data points in Fig. 9). This observation is attributed to the fact that the growth of hailstones beyond the embryo stage relies on the accretion of supercooled water rather than ice crystals (Lamb and Verlinde, 2011). That is,Consequently,

Formatted: Normal (Web), Justified, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted

Formatted

the hailstone recorded not only insoluble particles ~~in graupels possibly affected~~during the embryo formation ~~of ice crystals and subsequently affected~~, but also insoluble particle in the ~~formation of~~ hailstone ~~embryos. The~~growth zone throughout the hailstorm. As a result, the size ~~distributions~~distribution of particles within the entire hailstones may represent the distribution of insoluble particles in ~~eight hailstone embryos (BJ1, BJ2, GY1, GY2, BS, FS, YT, and GYA) were shown in Fig. 8.~~ deep convection regions where the hailstones went through.

Formatted

Formatted

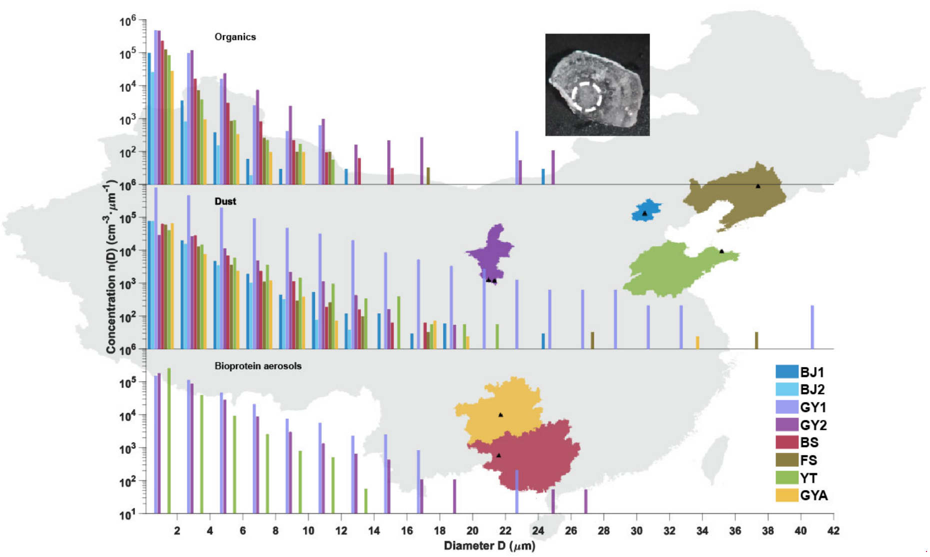
Formatted

Formatted

Formatted

Formatted

Formatted



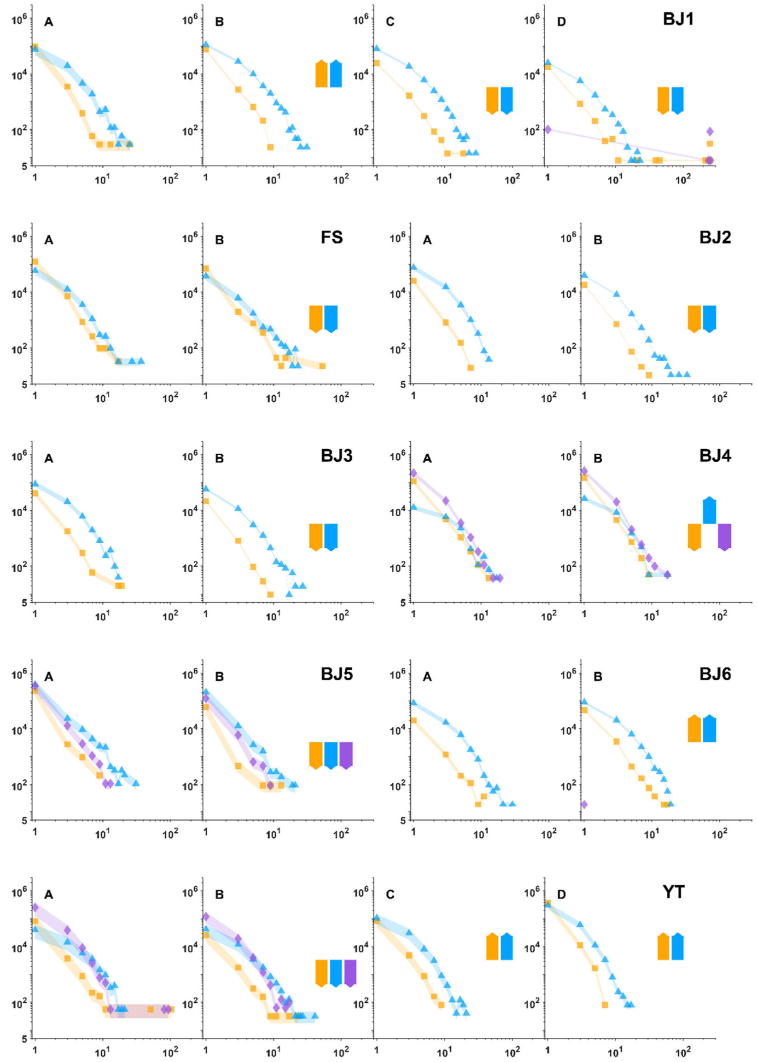


Fig. 8:9a Size distribution of insoluble particles in hailstone embryos. Provinces of China, from which the hailstones were collected, are shown in different colors. Black triangles indicate the locations of hailstone sample collection. The white dashed circle shows part of the hailstone embryo. Abbreviations (corresponding to Table. 1): BJ, Beijing City; GY, Guyuan City; BS, Baise City; FS, Fushun City; YT, Yantai City; GYA, Guiyang City.

Formatted: Font: +Body (等线), Not Bold

Formatted: Font: +Body (等线), Not Bold

Formatted: Font: +Body (等线), Not Bold

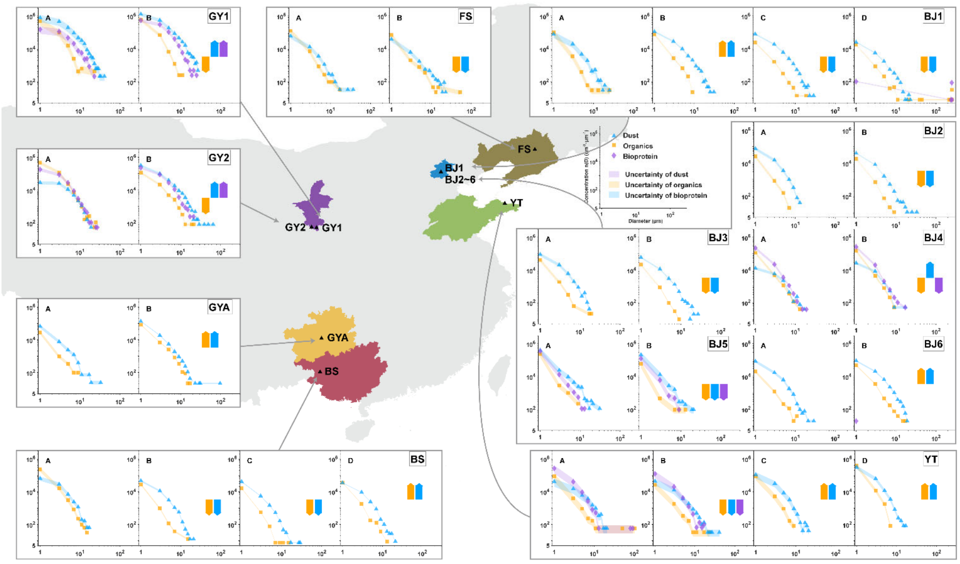
As mentioned above, BJ2 represented BJ2–BJ6. The variations in number concentrations of dust and bioprotein insoluble particles indicated that particle number concentrations decreased exponentially with particle diameter, with marked variation observed among hailstorms. The distribution distinguished organics from dust and bioprotein aerosols as the number concentrations of organics from all samples decreased with particle diameter before 8 μm, while those of GY1 and GY2 fluctuated starting at diameters of 8 μm and 12 μm, respectively. This was likely due to some uncontrolled residential and industrial coal burning in GY (Guyuan City). A great variance existed in size distribution of both organics and dust. The number concentrations of organics from a hailstone embryo were 1 to 390 times to those from different hailstone embryos at the same diameter. The number concentrations of dust from a hailstone embryo were 1 to 527 times to those from different hailstone embryos at the same diameter. The number concentrations of dust from BJ1, BJ2, and GY1 were at least 3 times higher than organics in particles of the same diameter in the range of 2–24 μm. Moreover, dust showed a wider size distribution than organics and bioproteins among all samples, since dust from GY1 had a higher number concentration and larger maximum size (42 μm) than from other hailstone embryos. Bioprotein aerosols, with high freezing efficiency, may have formed initial ice particles in GY1, GY2, and YT, while dust or organics caused initial ice particle formation in hailstorms in cases lacking bioprotein aerosols. All hailstone embryos contained organics and dust, but not all hailstone embryos contained a significant amount of bioprotein aerosols. There were uncertainties in quantification of biological aerosols, due to poor understanding of biological transport and transformation processes (Fröhlich-Nowoisky et al., 2016).

3.3 Variation in hailstone shells

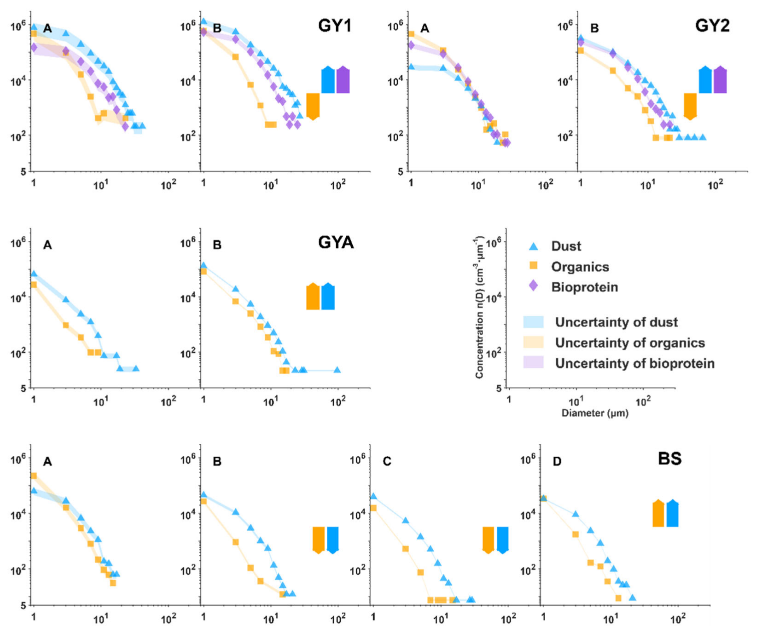
Size distribution of each species differed little in characteristics in outer shells with the embryos (Fig. 9). For a four-shell hailstone, the number concentrations of insoluble particles showed V-shaped (BS and YT) or inverse V-shaped (BJ1) distributions from embryo to crust. Five of nine two-shell hailstones showed higher number concentrations of dust in crusts than embryos, while seven of them showed higher number concentrations of organics in embryos than crusts. However, quantification of the differences in number concentration varied little among shells. The 90.5% points showed that differences in number concentration of the same kind particles in a shell compared to the previous shell at the same diameter was within twice, and the maximum differences was up to 9 times (294 data points in Fig. 9). This was because the growth of hailstones beyond the embryo stage depends on the accretion of supercooled water rather than ice crystals the (Lamb and Verlinde, 2011). The hailstone recorded not only insoluble particles when the embryo formed, but also insoluble particle in the hailstone growth zone throughout the hailstorm. Thus, the size distribution of particles within the whole hailstones may represent the distribution of insoluble particles in deep convection where the hailstones pass through.

Formatted

Formatted: Font: 宋体, 小五



~~Fig. 9: Size distribution of insoluble particles present in~~ natural shells of 12 hailstones~~. The diameter interval on the x-axis~~ is ~~2 µm. The y-axis shows the particle number concentration from *D* − 1 µm to *D* + 1 µm.~~represented. Blue triangles, orange ~~rectangles~~squares, and purple diamonds are used to indicate dust, organics, and bioprotein aerosols, respectively. The natural shells ~~were named~~are denoted alphabetically with capital letters (shell A refers to embryos~~,~~ and shell B/D refers to the crust of hailstones). The arrow direction ~~indicates~~illustrates the t~~r~~endency of particle number concentration in ~~this~~each layer ~~with regard~~compared to the previous ~~layer. Uncertainty~~shell. Shading is ~~indicated by shading. Calculations are described in detail in~~ employed to indicate uncertainty. Detailed calculations are provided in the supplementary information. Abbreviations (corresponding to Table~~.~~ 1): BJ~~, Beijing City; GY, Guyuan City~~ - BeiJing; BS~~, Baise City~~ - BaiSe; FS~~, Fushun City; YT, Yantai City~~ - FuShun; GY - GuYuan; GYA~~, Guiyang City~~ - GuiYAng; YT - YanTai.



~~**Fig. 9b is a continuation of Fig. 9a.**~~

3.4 Logarithmic normal distribution of dust and organics

The size distributions of dust and organics in the whole hailstone can be described by a logarithmic normal distribution (Fig. 10a) (Lamb and Verlinde, 2011):

$$\cancel{n(ln\, D) = \frac{N}{\sqrt{2\pi}\, ln\, \sigma_g} \cdot exp\left[-\frac{(ln\, D - ln\, r_g)^2}{2\, ln^2\, \sigma_g}\right], (D > 0.2\, \mu m)} \quad (12)$$

$$n(ln\, D) = \frac{N}{\sqrt{2\pi}\, ln\, \sigma_g} \cdot exp\left[-\frac{(ln\, D - ln\, D_g)^2}{2\, ln^2\, \sigma_g}\right], (D > 0.2\, \mu m) \quad (17)$$

Where $n(\ln D)$ is the number concentration of insoluble particles per cubic centimeter volume water ranging from $\ln D - \frac{1}{2}\mathrm{d}\ln D$ to $\ln D + \frac{1}{2}\mathrm{d}\ln D$. Here, D represents the diameter of particles (in micrometers), ~~$\ln r_g$~~ $ln\, D_g$ is the geometric mean diameter, and $\ln \sigma_g$ is the geometric standard deviation (Lamb and Verlinde, 2011). The number of bioprotein aerosols was below the limit of detection in some samples, so that, only the curves of organics and dust were fitted. The fitting parameters of the same species were aggregated in parameter space, and were suspected to be related to the physical properties of each species, requiring further studies for confirmation. Moreover, the fitting parameters of organics and dust particles were

Formatted: Line spacing: single, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Formatted

Formatted: Font: 小五, Bold

Formatted: Normal, Left, Line spacing: single clustered into two centroids (Fig. 10b) by the K-means method, which indicated that organics and dust have two classic modes (classic mode of organics: ~~$\ln r_o$~~ $\ln D_o$ = -0.70 ~~μm~~, $\ln \sigma_o$ = 0.91 ~~μm~~, and N_o = 9.19 × 10^5 cm^{-3}; classic mode of dust: ~~$\ln r_d$~~ $\ln D_d$ = 0.11 ~~μm~~, $\ln \sigma_d$ = 1.07 ~~μm~~, and N_o = 1.~~58~~59 × 10^6 cm^{-3}). That is, insoluble organics in hailstones are usually smaller in diameter and present in lower amounts than dust. Regardless of fine or coarse particles (D < 0.5 μm in diameter were not considered in reference to DeMott et al. (DeMott et al., 2010)), the number concentration of dust was up to 2 orders of magnitude higher than the number concentration of organics. These observations indicated that dust accounted for the major portion of particles in eight hailstorms (no considering about bioprotein), which was consistent with the observations of embryos described above.

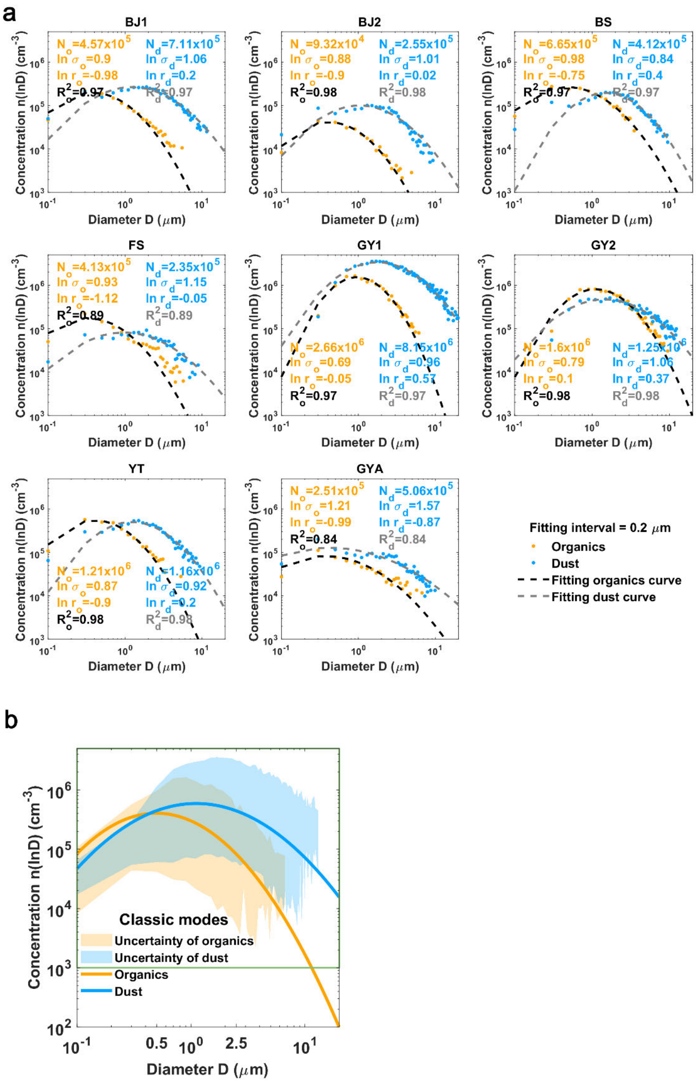


Fig. 10: Fitting size distribution functions of organics and dust contained in the whole hailstone. (a) Fitting parameters of logarithmic normal distributions of BJ1, BJ2, BS, FS, GY1, GY2, YT, GYA. (b) Classic modes of dust and organics (interval of data is 0.2 μm and fitting curves painted with interval of 0.02 μm). The fitting range of (a) is shown with a green rectangle. The centroid of the organics fitting parameter (orange line) is $\ln\sigma_o$ = 0.91 ~~μm, $\ln r_o$~~ $\ln D_o$ = −0.70 ~~μm~~, and N_o = 9.19 × 10^5 cm^{-3}. The centroid of the dust fitting parameter (blue line) is $\ln\sigma_d$ = 1.07 ~~μm,~~ ~~$\ln r_d$~~ $\ln D_d$ = 0.11 ~~μm~~, and N_d = 1.59 × 10^6 cm^{-3}. Shading showed uncertainty of organics and dust. Abbreviations (corresponding to Table 1): BJ~~, Beijing City; GY, Guyuan City~~ - BeiJing; BS~~, Baise City~~ - BaiSe; FS~~, Fushun City; YT, Yantai City~~ - FuShun; GY - GuYuan; GYA~~, Guiyang City.~~

- GuiYAng; YT - YanTai.

4 Conclusions

This was the first study to simultaneously analyze both the number concentrations and species (organics, dust and bioproteins) of insoluble particles in hailstones. Analysis of insoluble particles present in hailstones, which participate in heterogeneous nucleating process as ice-nucleating particles in a deep convection(Lamb and Verlinde, 2011), provides a new approach for refinement of particle observation in severe storms and the understanding of hailstone formation.including organics, dust and bioproteins) of insoluble particles in hailstones. The findings from this analysis offer valuable insights into particle observations within severe storms. Understanding the number concentration and composition of these insoluble particles is crucial, as they play a significant role as ice-nucleating particles during the heterogeneous nucleation process in deep convection.

The size distribution of insoluble particles in hailstones from the same hailstorm showed less variation than those from different hailstorms. One possible reason is that updrafts of hailstorms brought insoluble particles from local surfaces or boundary layers into deep convective clouds. Moreover, part of these insoluble particles participate in freezing initial ice particles to form one type of hailstone embryos. Almost all insoluble particles in hailstone embryos analyzed in this study showed an exponential size distribution, which was consistent with the effects of gravity. The number concentrations of organics and dust from different hailstone embryos differed up to 389 times and 526 times at the same diameter, respectively. Moreover, almost all insoluble particles in hailstone embryos analyzed in this study showed an exponential size distribution, which was consistent with the effects of gravity. The number concentrations of organics and dust from different hailstone embryos differed up to 389 times and 526 times at the same diameter, respectively. The changes in paticle concentration may lead to at leat one-order-of-magnitude viarance in ice-nucleating particle (DeMott et al., 2010). Additionally, size distribution of insoluble particles varied in shells up to 27Hailstone samples with high insoluble particle content, i.e., GY1 and GY2, showed significantly lower total column water vapor values and smaller depth between freezing level height and orography within one hour before hailstorm occurrence, compared to other samples (Hersbach et al., 2018). The competition of condensation and shorter updraft pathway might be responsible for the high number concentrations of organics, dust, and bioproteins in GY1 and GY2. Size distribution of insoluble particles varied in shells up to 9 times, which was much small than differences with different hailstorms.

Two classic logarithmic normal distribution models were applied to fit the size distribution modes of organics and dust within hailstones were fitted as logarithmic normal distribution for, providing a description of insoluble particles in the deep convection where the hailstones grew up. Theduring hailstone formation. The analysis of the two classic size distribution modes of insoluble particles suggested thatindicated a significant presence of dust occupied the major fraction, without

~~taking~~considering bioprotein ~~into account. Besides, there is~~. Furthermore, a positive correlation exists between the number concentrations of insoluble particles and that of ice-nucleating particles in hailstones, specifically for corresponding species (Ren et al., 2023, submitted, figure not shown). ~~Further~~A further measurement of ice-nucleating particles by drop-freezing experiments will establish the relationship between insoluble particles and immersion ice-nucleating particles. Combination of these results with future experiments to determine the number concentrations and species of particles from local observations will establish the relationship between surface observation and ice-nucleating particles in deep convective clouds, which will lead to improvement of the parameterization of ice-nucleating particles in both weather and climate models.

~~However~~Nonetheless, two kinds of classic size distribution modes of organics and dust in hailstones were performed, but a more robust classic mode required a larger number of samples. In future, for ~~any~~ climate or weather models, the classic mode can be assumed as the mean state to describe the characteristics of insoluble particles in supercooling water. In addition, this study did not attempt to parameterize bioprotein aerosols, because there was a great uncertainty in quantification due to poor understanding of biological processes (Fröhlich-Nowoisky et al., 2016). Further collaborative studies are required to gain a better understanding of biological processes to establish the classic bioprotein mode.

Code availability

Self-organized maps algorithm is functions on MATLAB

https://ww2.mathworks.cn/help/deeplearning/ref/selforgmap.html

Random forest algorithm is functions on MATLAB

https://ww2.mathworks.cn/help/stats/treebagger.html?searchHighlight=TreeBagger&s_tid=srchtitle_TreeBagger_1

The 10-fold stratified cross-validation algorithm is functions on MATLAB

https://ww2.mathworks.cn/help/stats/cvpartition.html?searchHighlight=cvpartition&s_tid=srchtitle_cvpartition_1

Identification algorithms are coded on MATLAB and will be made available on request.

Data availability

Data will be made available on request.

Author contributions

Haifan Zhang wrote the original draft under the concept presented by Qinghong Zhang. Haifan Zhang, Xiangyu Lin and Chan-Pang Ng participated in preprocess and reservation of hailstones from volunteers. Haifan Zhang and Xiangyu Lin sliced hailstones using machine manufactured by Kai Bi and performed the experiments on analyzing element weight ratio of

Formatted
Formatted
Formatted
Formatted
Formatted

insoluble particles with help of Li Chen. Kai Bi also provided hailstones BJ2 ~ BJ6. Machine learning on identification of particles is operated by Haifan Zhang. Yangze Ren and Huiwen Xue compared ice nucleation particles from drop-freezing experiments with our data. Zhuolin Chang provided hailstones GY1 and GY2. All authors discussed and contributed to the final manuscript. Qinghong Zhang directed this project.

Competing interests

The authors declare no competing interests.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant Nos. 42030607 and 41930968), the Innovation Project of the China Meteorological Administration (Grant No. CXFZ2021J038) and the Key R & D projects in Ningxia Hui Autonomous Region (2022BEG02010). The authors thank Cai Yao from the Meteorological Bureau of Guangxi, China in collecting hailstones BS in Guangxi. The authors thank volunteers in collecting hailstones. The authors thank Prof. Jiwen Fan from Pacific Northwest National Laboratory of the United States for discussions.

References:

Aztec User Manual: https://utw10193.utweb.utexas.edu/InstrumentManuals/Oxford%20EDS%20AZtec%20User%20Manual.pdf, last access: 22 August 2023.

~~Ault, A. P., Peters, T. M., Sawvel, E. J., Casuccio, G. S., Willis, R. D., Norris, G. A., and Grassian, V. H.: Single-Particle SEM-EDX Analysis of Iron-Containing Coarse Particulate Matter in an Urban Environment: Sources and Distribution of Iron within Cleveland, Ohio, Environ. Sci. Technol., 46, 4331–4339, https://doi.org/10.1021/es204006k, 2012.~~

Battaglia, A., Mroz, K., and Cecil, D.: Satellite hail detection, in: Precipitation Science, Elsevier, 257–286, https://doi.org/10.1016/B978-0-12-822973-6.00006-8, 2022.

Beal, A., Martins, J. A., Rudke, A. P., de Almeida, D. S., da Silva, I., Sobrinho, O. M., de Fátima Andrade, M., Tarley, C. R. T., and Martins, L. D.: Chemical characterization of PM2.5 from region highly impacted by hailstorms in South America, Environ. Sci. Pollut. Res., 29, 5840–5851, https://doi.org/10.1007/s11356-021-15952-6, 2022.

Calinski, T. and Harabasz, J.: A dendrite method for cluster analysis, Commun. Stat. - Theory Methods, 3, 1–27, https://doi.org/10.1080/03610927408827101, 1974.

Chen, Q., Yin, Y., Jiang, H., Chu, Z., Xue, L., Shi, R., Zhang, X., and Chen, J.: The Roles of Mineral Dust as Cloud Condensation Nuclei and Ice Nuclei During the Evolution of a Hail Storm, J. Geophys. Res. Atmos., 124, 14262–14284, https://doi.org/10.1029/2019JD031403, 2019.

Davies, D. L. and Bouldin, D. W.: A Cluster Separation Measure, IEEE Trans. Pattern Anal. Mach. Intell., PAMI-1, 224–227, https://doi.org/10.1109/TPAMI.1979.4766909, 1979.

DeMott, P. J., Prenni, A. J., Liu, X., Kreidenweis, S. M., Petters, M. D., Twohy, C. H., Richardson, M. S., Eidhammer, T., and Rogers, D. C.: Predicting global atmospheric ice nuclei distributions and their impacts on climate, Proc. Natl. Acad. Sci., 107, 11217–11222, https://doi.org/10.1073/pnas.0910818107, 2010.

DeMott, P. J., Prenni, A. J., McMeeking, G. R., Sullivan, R. C., Petters, M. D., Tobo, Y., Niemand, M., Möhler, O., Snider, J. R., Wang, Z., and Kreidenweis, S. M.: Integrating laboratory and field data to quantify the immersion freezing ice nucleation activity of mineral dust particles, Atmos. Chem. Phys., 15, 393–409, https://doi.org/10.5194/acp-15-393-2015, 2015.

Fröhlich-Nowoisky, J., Kampf, C. J., Weber, B., Huffman, J. A., Pöhlker, C., Andreae, M. O., Lang-Yona, N., Burrows, S. M., Gunthe, S. S., Elbert, W., Su, H., Hoor, P., Thines, E., Hoffmann, T., Després, V. R., and Pöschl, U.: Bioaerosols in the Earth system: Climate, health, and ecosystem interactions, Atmos. Res., 182, 346–376, https://doi.org/10.1016/j.atmosres.2016.07.018, 2016.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1959 to present, Copernicus Clim. Chang. Serv. Clim. Data Store (CDS).[data set], https://doi.org/https://doi.org/10.24381/cds.adbb2d47, 2018.

Hoose, C. and Möhler, O.: Heterogeneous ice nucleation on atmospheric aerosols: a review of results from laboratory experiments, Atmos. Chem. Phys., 12, 9817–9854, https://doi.org/10.5194/acp-12-9817-2012, 2012.

~~Hoose, C., Kristjánsson, J. E., Chen, J.-P., and Hazra, A.: A Classical-Theory-Based Parameterization of Heterogeneous Ice Nucleation by Mineral Dust, Soot, and Biological Particles in a Global Climate Model, J. Atmos. Sci., 67, 2483–2503, https://doi.org/10.1175/2010JAS3425.1, 2010.~~

~~Jouzel, J., Merlivat, L., and Roth, E.: Isotopic study of hail, J. Geophys. Res., 80, 5015–5030, https://doi.org/10.1029/JC080i036p05015, 1975.~~

~~Kirpes, R. M., Bondy, A. L., Bonanno, D., Moffet, R. C., Wang, B., Laskin, A., Ault, A. P., and Pratt, K. A.: Secondary sulfate is internally mixed with sea spray aerosol and organic aerosol in the winter Arctic, Atmos. Chem. Phys., 18, 3937–3949, https://doi.org/10.5194/acp-18-3937-2018, 2018.~~

Knight, N. C.: The Climatology of Hailstone Embryos, J. Appl. Meteorol., 20, 750–755, https://doi.org/10.1175/1520-0450(1981)020<0750:TCOHE>2.0.CO;2, 1981.

Kohonen, T.: The self-organizing map, Proc. IEEE, 78, 1464–1480, https://doi.org/10.1109/5.58325, 1990.

Lamb, D. and Verlinde, J.: Physics and Chemistry of Clouds, Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9780511976377, 2011.

Li, X., Zhang, Q., Zhu, T., Li, Z., Lin, J., and Zou, T.: Water-soluble ions in hailstones in northern and southwestern China, Sci. Bull., 63, 1177–1179, https://doi.org/10.1016/j.scib.2018.07.021, 2018.

Li, X., Zhang, Q., Zhou, L., and An, Y.: Chemical composition of a hailstone: evidence for tracking hailstone trajectory in deep convection, Sci. Bull., 65, 1337–1339, https://doi.org/10.1016/j.scib.2020.04.034, 2020.

Michaud, A. B., Dore, J. E., Leslie, D., Lyons, W. B., Sands, D. C., and Priscu, J. C.: Biological ice nucleation initiates hailstone formation, J. Geophys. Res. Atmos., 119, 12,186-12,197, https://doi.org/10.1002/2014JD022004, 2014.

Prenni, A. J., Demott, P. J., Rogers, D. C., Kreidenweis, S. M., Mcfarquhar, G. M., Zhang, G., and Poellot, M. R.: Ice nuclei characteristics from M-PACE and their relation to ice formation in clouds, Tellus B, 61, 436–448, https://doi.org/10.1111/j.1600-0889.2009.00415.x, 2009.

Rogers, D. C., DeMott, P. J., Kreidenweis, S. M., and Chen, Y.: A Continuous-Flow Diffusion Chamber for Airborne Measurements of Ice Nuclei, J. Atmos. Ocean. Technol., 18, 725–741, https://doi.org/10.1175/1520-0426(2001)018<0725:ACFDCF>2.0.CO;2, 2001.

Rosinski, J.: Solid Water-Insoluble Particles in Hailstones and Their Geophysical Significance, J. Appl. Meteorol., 5, 481–

492, https://doi.org/10.1175/1520-0450(1966)005<0481:SWIPIH>2.0.CO;2, 1966.

Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math., 20, 53–65, https://doi.org/10.1016/0377-0427(87)90125-7, 1987.

Sibson, R. and Hartigan, J. A.: Clustering Algorithms., Appl. Stat., 25, 70, https://doi.org/10.2307/2346526, 1976.

Tao, J., Zhang, L., Cao, J., and Zhang, R.: A review of current knowledge concerning PM2.5 chemical composition, aerosol optical properties and their relationships across China, Atmos. Chem. Phys., 17, 9485–9518, https://doi.org/10.5194/acp-17-9485-2017, 2017.

Taylor, J. R.: An Introduction to Error Analysis, Second edi., University Science Books, 330 pp., 1997.

Vali, G.: Ice Nucleation Relevant to Formation of Hail, 1968.

Vergara-Temprado, J., Miltenberger, A. K., Furtado, K., Grosvenor, D. P., Shipway, B. J., Hill, A. A., Wilkinson, J. M., Field, P. R., Murray, B. J., and Carslaw, K. S.: Strong control of Southern Ocean cloud reflectivity by ice-nucleating particles, Proc. Natl. Acad. Sci., 115, 2687–2692, https://doi.org/10.1073/pnas.1721627115, 2018.

Formatted: Font: 10 pt