

# Climate impacts and future trends of hailstorms in China based on millennial records

Received: 18 December 2024

Accepted: 7 August 2025

Published online: 02 September 2025

Qinghong Zhang<sup>1,5</sup>✉, Rumeng Li<sup>1,4,5</sup>, Wenhong Li<sup>2</sup>✉, Xiaofei Li<sup>3</sup>,  
Chan-Pang Ng<sup>1</sup>, Shiyi Zhang<sup>1</sup> & Chuanfeng Zhao<sup>1</sup>

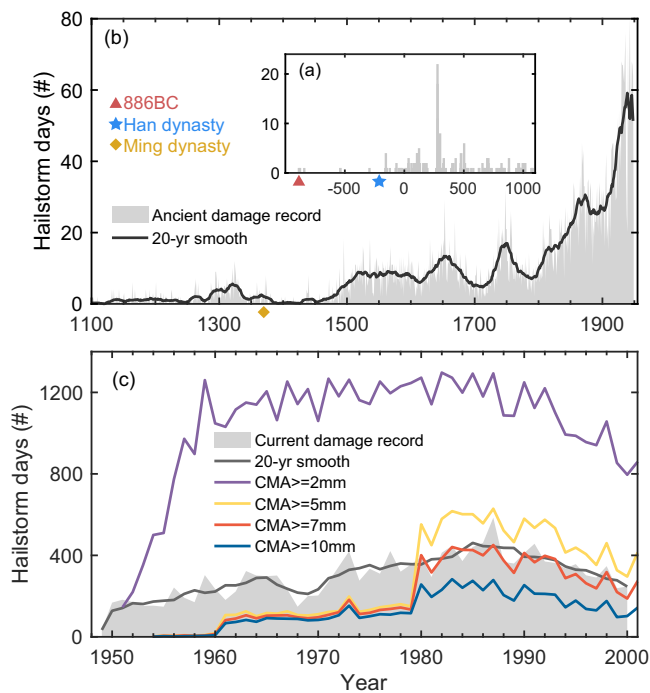
Understanding how hailstorm trends have changed in the context of climate change is a persistent challenge, mainly because of the lack of long-term consistent observations of hailstorms. Here, we leverage hail damage records from Chinese historical books and extend hailstorm records to approximately 2890 years ago, exploring variations in the number of hailstorm days between 1500 and 1949 based on reliable and consistent data. We show that the number of hailstorm days was constant before 1850, but has increased significantly afterwards. This increase in hailstorm days seems to be associated with the increase in surface temperature after the population effect is removed. In addition to the trend, hailstorm activity is found to display both quasi-centennial and multidecadal variability, with the former (later) dominating before (after) the 1850s, driven by the Pacific Decadal Oscillation (PDO). These results suggest that long-term changes in hailstorm days in China are modulated by climate warming and natural variability, via the PDO. Future projections based on different climate change scenarios and a convolutional neural network model show a further increase in the number of hailstorm days in the 21<sup>st</sup> century.

Hailstorms, among the most intense deep convection systems on Earth, play important roles in the hydrological cycle and general circulation of the atmosphere. Hailstones formed in hailstorms can be damaging and impact society in many areas of the world<sup>1</sup>. Therefore, understanding how and why hailstorm occurrence has varied in the past and predicting its variations/changes in the future are crucial to disaster management. Recently, the changes in hailstorm trends associated with anthropogenic climate warming have drawn considerable attention from research communities worldwide<sup>2</sup>. The main challenges are the inadequacies in monitoring systems from both spatial and temporal perspectives for such small-scale weather phenomena and the lack of knowledge regarding the microphysics of local-scale deep convection and the link between

thermal dynamics and microphysics<sup>3,4</sup>. Therefore, changes in hailstorm occurrence in the context of a warming climate were considered as being low confident in both IPCC-AR5 to IPCC-AR6 reports<sup>4,5</sup>.

Hailstorm observations in the past have displayed considerable heterogeneity, which makes studying historical hail trends from a global perspective exceedingly challenging. In a particular region, hail observations from various sources may show different trends. For example, observations of United States hailstorms from public reports revealed a significant increase in frequency over the latter half of the twentieth century, whereas data from meteorological stations revealed no significant trend<sup>6,7</sup>. The artificial increase in reporting is thought to be caused by the growth of the population<sup>7</sup>.

<sup>1</sup>Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871, China. <sup>2</sup>Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, NC, USA. <sup>3</sup>College of Urban and Environmental Sciences, Northwest University, 710027 Xi'an, China. <sup>4</sup>Present address: Institute of Meteorology and Climate Research (IMK-TRO), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. <sup>5</sup>These authors contributed equally: Qinghong Zhang, Rumeng Li. ✉ e-mail: [qzhang@pku.edu.cn](mailto:qzhang@pku.edu.cn); [wenhong.li@duke.edu](mailto:wenhong.li@duke.edu)



**Fig. 1 | Long-term records of hailstorm days in China.** **a** Ancient hail damage records from 886 BC to 1100 (grey bars), with key historical markers at 886 BC and the Han dynasty indicated by pink triangle and blue star, respectively. **(b)** Ancient damage records from 1100–1948 (grey bars); the black solid line shows the 20 year smoothed time series. The beginning of the Ming dynasty is marked by a yellow diamond. **c** Hail damage data from 1949–2000 (grey shading); colored lines represent CMA (China Meteorological Administration) station records for hailstorms with maximum hailstone sizes  $\geq 2$  mm (purple),  $\geq 5$  mm (yellow),  $\geq 7$  mm (orange), and  $\geq 10$  mm (blue). The grey solid line shows the 20 year smoothed hail damage data.

Although meteorological station observations are continuous and coherent in several countries in eastern Asia and a robust decrease in hailstorm days and occurrence has been reported over the past several decades<sup>8–15</sup>, the coarse spacing distance of the meteorological stations has resulted in a lack of spatial representativeness. Most importantly, these observations have been limited in the last half century, and the data period is short to reveal the variations/changes in hailstorms with near-term climate warming since the Industrial Revolution.

China is a country with a long history of >5000 years. The historic records in China span to the Yin dynasty 3000 years ago<sup>16</sup>. There are many dynastic histories, local chronicles, agricultural records, geographic monographs, personal travelogues, and other archival materials available. Natural disasters, such as floods, earthquakes, and hail damage, have made significant social impacts on people, and have therefore been recorded<sup>16,17</sup>. The very first hailstorm damage record occurred in 886 BC (Fig. 1a); however, regular local chronicles started in the Han dynasty at approximately 200 BC. The valuable, long-term hail damage records provide us with an opportunity to extend the analysis of hailstorm occurrence to 200 BC. In this work, we introduce a long-term hail damage record dataset and analyse the variations in hailstorm occurrence after 1500, covering the periods both pre- and post-Industrial revolution, as data after 1500 are thought to be more reliable than earlier records. We also project the future occurrence of hailstorm days at the yearly scale via artificial intelligence technology under different climate change scenarios (RCP 2.6, RCP 4.5, RCP 6.0 and RCP 8.5 from the CMIP5 multimodel ensemble mean) at the yearly scale with the use of a convolutional neural network model.

## Results

### Variations in hailstorm occurrence

Figure 1 shows the interannual variations in the number of hailstorm days derived from ancient damage records from 886 BC–1949 AD (bars in Fig. 1a, b), current damage records from 1949–2000 (grey shading in Fig. 1c), and meteorological station records from 1949–2000 (coloured solid lines in Fig. 1c). The territory of China changed significantly prior to the Ming dynasty (ruled China beginning in 1368 AD<sup>18</sup>). Before 1500 AD, the annual number of hailstorm damage days was typically fewer than 10, and the hail damage records were usually discontinuous (Fig. 1a, b). While after 1500 AD, the hail damage records were relatively reliable, considering territory of China change little after Ming dynasty. Following 1949, the hailstorm days identified from hail damage records (grey shading and grey solid line in Fig. 1c) show good agreement with meteorological station observation. This alignment is particularly evident for hailstone with a diameter of 7 mm or larger (orange solid line in Fig. 1c) after 1979<sup>13</sup>. These findings suggest that the use of hail damage data to study hailstorm activity is quite reliable, which is consistent with the findings of previous studies<sup>17,19</sup>. Considering that ancient and current hail damage datasets were recorded in different systems, we focus on analysing the features of hailstorm damage days on the basis of ancient hail damage data from 1500 to 1948.

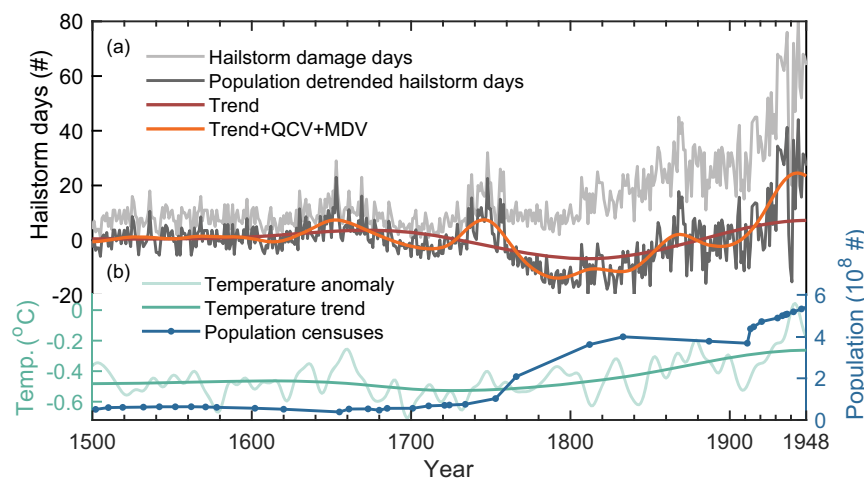
The annual variations in hailstorm damage days during this period (Figs. 1b, 2) display two characteristics. First, there was a significant increasing trend at the beginning of the nineteenth century. Specifically, the number of hailstorm damage days increased from 10 days per year in the 1700s to 60 days per year in the 1940s, an average rate of increase of 3.9 days per decade. The increase during the 1900s–1940s, reaching an average rate of 8.3 days per decade, more than doubled that during the 1850s–1890s (Fig. 2). Second, multidecadal and quasicentennial variations were obvious. A power spectra analysis of hailstorm damage days suggested that a 90 year cycle is the dominant frequency, and this result is significant considering a red-noise background at the 95% upper confidence interval; moreover, a 56 year cycle can be identified as the second-most-dominant frequency (Fig. S4).

Previous studies have demonstrated that population growth may be associated with the increase in the yearly number of hail reports in the United States, thus increasing hailstorm days<sup>7,20</sup>. To eliminate the possible artificial effects of population growth on the number of records, the population effect were first processed via a regression method<sup>21,22</sup> (for more details, see the Methods). Figure 2 shows the population-detrended number of hailstorm days (referred to as number of hailstorm days in the following text; dark grey solid line), which are used in the following analysis.

In general, long-term variations in the number of hailstorm days after population effects are removed still show a clear increase after the 1850s. In addition, the power spectra analysis reveals two cycles at the quasicentennial and multidecadal scales (orange line in Fig. S4). These results suggest that hailstorm activity exhibits quasicentennial and multidecadal cycles, in addition to an increasing trend after 1500. It is interesting to explore what may have caused such changes/variations in hailstorm days.

### Influence of temperature changes

Current knowledge suggests that areas of high convective available potential energy (CAPE), with high energy level for prevailing circulation; abundant moisture, which provides supercooled water for hailstone growth; and strong vertical wind shear, which contributes to convective processes<sup>23,24</sup> (although not necessarily simultaneously), are favourable environments for hailstorm formation. Since long-term records of wind are not available, we show the possible influence of surface temperature changes, which are related to convective instability and atmospheric moisture, the first two factors listed above, on hailstorm occurrence in China.



**Fig. 2 | Long-term variations in hailstorm damage days, population, and temperature from 1500 to 1948. a** Observed hailstorm damage days (light grey greysolid line) and population-detrended hailstorm damage days (dark grey solid line). The brown line represents the nonlinear trend component of population-detrended hailstorm damage days decomposed via the CEEMDAN method. The

orange line shows the nonlinear trend, quasicentennial variability (QCV), and multi-decadal variability (MDV) components. **b** Time series of the population (blue dotted line), temperature anomaly (light green line), and the nonlinear temperature trend component (dark green line), which is also derived from the CEEMDAN method.

Five long-term temperature datasets<sup>24–28</sup> are widely used by the climate community<sup>5</sup>. These five datasets differ slightly dating back to 1500 (Fig. S2). As the temperature in the global Mann dataset has the highest correlation coefficient with the number of hailstorm days, the following discussion is based on the Mann dataset (Fig. 2).

Figure 2 shows that the global mean temperature changed only slightly before 1850; it increased by approximately 0.8 degrees from 1850–1948, and most of the temperature increase occurred in the last 50 years of that period. Correspondingly, the number of hailstorm days increased from 20–60 days per year from 1850–1948, and most of the increase occurred in the last 50 years of the period (Fig. 2).

To quantify the influence of the global mean surface temperature on the variations in the number of hailstorm days in China, we further used the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method to obtain the features (Figs. S5 and S6), such as different timescale variabilities and/or nonlinear trends, of hailstorm days and temperature. Notably, CEEMDAN can be used to decompose the time series of a variable into different time scales without any prior subjective criteria<sup>29–31</sup>. The trend of hailstorm days is significantly correlated with that of the global mean temperature (Mann dataset), with a coefficient of 0.99 after 1800, indicating a high possibility that the increase in the number of hailstorm days was caused by an increased global temperature.

In addition to long-term trends, variabilities at different timescale were observed for the number of hailstorm days, implying that other climate factors may contribute to changes in hailstorm trends. As this research focuses on variations in the number of hailstorm days at long time scales, determining the climate factors that dominate the variations in hailstorm days at decadal and quasicentennial scales is important.

### Influence of the PDO

Previous studies have suggested that multidecadal-scale precipitation and surface temperature in China are largely impacted by the Pacific Decadal Oscillation (PDO) and North Atlantic Oscillation (NAO)<sup>32–40</sup>. To further assess the possible influences of the PDO and NAO on hailstorm trends, we used the CEEMDAN method to analyse their variations at different time scales. (Fig. S7–S8). At the quasicentennial scale, there is a significant negative correlation between the PDO and the number of hailstorm days, especially before 1850. At the multidecadal

scale, the correlation between the PDO and number of hailstorm days is positive at the 95% confidence level (Fig. 3). Furthermore, the correlation between the PDO and number of hailstorm days is significant, with a coefficient of  $-0.96$  ( $0.57$ ) for the quasicentennial variation (multidecadal variation) if an 80–100 (50–70) year fast Fourier transform bandpass filter is applied (Fig. S9b). However, the NAO is not significantly correlated with the number of hailstorm days at the quasicentennial scale or the multidecadal scale (Fig. 3).

Previous research indicated that quasicentennial and multidecadal oscillations were dominant in the PDO before and after the 1850s, respectively<sup>41</sup>. The quasicentennial cycle corresponds to the solar irradiance fluctuation (the Gleissberg cycle), and the multidecadal oscillation tends to be associated with global warming<sup>41</sup>. However, we cannot rule out the effects of different observational sources between the preindustrial and post-industrial periods as well as the low confidence in decadal PDO records prior to 1950<sup>4</sup>.

The different relationships between the PDO and the number of hailstorm days at different time scales may be associated with different atmospheric circulations and climatic backgrounds during these periods. During the Little Ice Age (LIA; 1450–1850), when quasicentennial variations dominated both the PDO and hailstorm trends, widespread negative precipitation anomalies were found from southern to northern China when the PDO was in a positive phase<sup>39</sup>. Previous studies have demonstrated that northerly wind anomalies dominate eastern China when the PDO is in the positive phase, resulting in a relatively dry and cold environment that inhibits precipitation<sup>34,39,40</sup>, which is unfavourable for hailstorm occurrence because of the relatively low CAPE and limited amount of precipitable water (PW). Thus, the correlation between the PDO and the number of hailstorm days was negative (Fig. S10). After 1850, with increasing temperature, multidecadal variations with the PDO became dominant<sup>41</sup>; during the positive PDO phase, abnormal southerly winds transported moisture to eastern China<sup>42</sup>, promoting the occurrence of hailstorms due to relatively high CAPE and PW levels (Fig. S10). Our analysis also revealed variations in the location and intensity of the PDO before and after 1850, which may deserve further analysis in the future.

### Future projections

Given that the nonlinear trend and quasicentennial and multidecadal variations in hailstorm days are influenced mainly by warming and the PDO, a one-dimensional convolutional neural network (CNN)

model<sup>43–46</sup> was developed to project the number of hailstorm days in the future. We chose the period of 1500–1948 as the training period and 1950–2099 as the prediction period. Trend and oscillation predictions were conducted separately. The model reasonably captures the changes in hailstorm trends between the 1950s and 1970s and between the 1980s and 2000s, respectively, when the PDO was in negative and positive phases (Fig. 4). This finding further suggests that the CNN model provides an accurate projection of the number of hailstorm days in China.

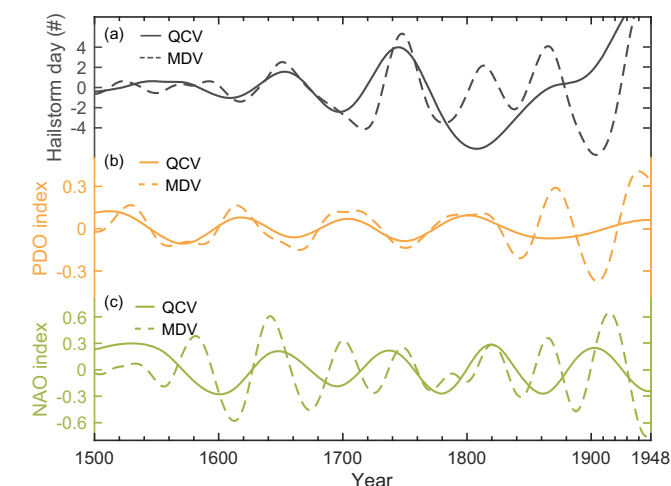
In the future, the global mean surface temperature is projected to increase by 0.8, 1.7, 2.2, and 4.0 °C under RCP 2.6, RCP 4.5, RCP 6.0 and RCP 8.5, respectively (green lines in Fig. 4). Models in the Coupled Model Intercomparison Project phase 5 (CMIP5) project a high possibility of a positive PDO from 2040 to 2080<sup>47</sup>. The number of hailstorm days in China is expected to increase substantially from now to 2060 s

(Fig. 4). Specifically, this number would increase to 620 days in 2072 from 300 days in 2017 under the RCP 8.5 scenario, an increase of 107%. The increases would be 76%, 75%, and 59% under the RCP 6.0, RCP 4.5 and RCP 2.6 scenarios, respectively. The CNN model also projects a decrease in the number of hailstorm days after the 2070s, presumably because of the PDO phase change. A sensitivity test using a synthetic PDO dataset in which the positive phase was manually extended until 2100 has been conducted. The results show that without a PDO phase change after 2070s, the predicted decrease in hailstorm days does not occur. The number of hailstorm days would never return to the value of 241 per year observed in the 20th century.

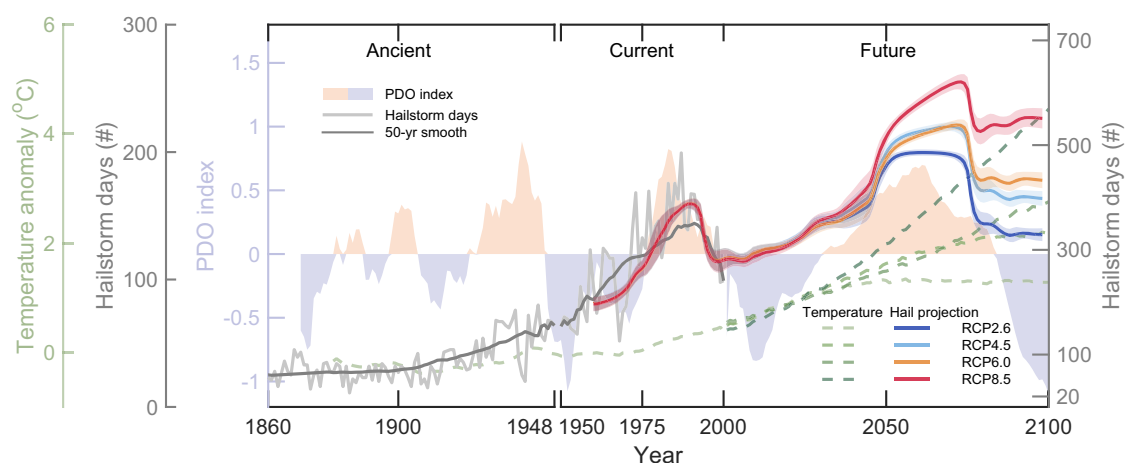
## Discussion

Studies of changes in hailstorm days in the context of climate warming are largely limited by the lack of observations. In China, reliable meteorological observation stations have been established since the middle of the last century, a period not sufficient for evaluating long-term responses. The long-term hailstorm damage records from ancient China thus provide us with valuable information for analyses. Notably, analyses of long-term hailstorm damage data suggested that the number of hailstorm days increased by approximately 3.9 days per decade in China with warming from 1850–1948. In addition, the PDO modulated regional hailstorm activity at the quasicentennial (80–100 years) and multidecadal (50–70 years) scales before and after 1850, respectively. A CNN model setup based on hailstorm damage data from 1500–1948 projected that hailstorm days are expected to increase as warming continues, largely in the next 40–50 years. Our results show that the number of hailstorm days at the end of this century will not return to the levels observed in the 20th century. Therefore, the challenge of disaster prevention related to hailstorms will increase as the 21<sup>st</sup> century continues.

Uncertainty remains in our study due to the exclusion of factors such as urbanization, aerosols, and land-use changes. Urbanization can increase surface temperatures and atmospheric instability, potentially intensifying updrafts necessary for hail formation. Aerosols influence cloud microphysics, may either suppress or invigorate hailstorms depending on their type and concentration. Land-use changes, such as deforestation or agriculture expansion, modify surface moisture and temperature gradients, thereby altering atmospheric instability and storm dynamics. Overall, these anthropogenic factors likely play a role



**Fig. 3 | Variability in quasicentennial and multidecadal components of hailstorm day and climate indices.** Solid lines represent the quasicentennial variability (QCV) components, and dashed lines represent the multidecadal variability (MDV) components. **a** Hailstorm day variability, **b** Pacific Decadal Oscillation (PDO) variability, **c** North Atlantic Oscillation (NAO) variability.



**Fig. 4 | Historical and projected trends of hailstorm days in relation to temperature anomalies and the PDO.** Annual hailstorm days are shown as light grey lines, with 50 year smoothed trends in dark grey for historical (pre-1950, left Y-axis) and current (1950–2000, right Y-axis) periods. Colored solid lines from 1960–2000 represent projections (post-2000, right Y-axis) based on observed temperature anomalies and the Pacific Decadal Oscillation (PDO) index. Colored solid lines from 2000–2100 show future projections under RCP scenarios: RCP-8.5 (red), RCP-6.0

(orange), RCP-4.5 (light blue), and RCP-2.6 (dark blue). Shading around projections indicates  $\pm 1$  standard deviation from 200 ensemble training runs. PDO index is shown as shading, with data from 1870 to 2013 based on HadISST1.1, and from 2014 onward derived from CMIP5 models<sup>47</sup>. Temperature anomalies are shown as dashed green lines (10 year low-pass filtered), with different shades representing the four RCP scenarios. Data from 1880 and 2000 are from NASA/GISS, and from 2000 to 2100 are from the CMIP5 multimodel ensemble mean.



in hailstorm occurrence, though their effects are highly dependent on local environmental contexts and feedback mechanisms. In addition, advances in communication technology and changes of observation systems over time may have influenced the frequency of recorded hail events<sup>48,49</sup>. Further research should explore how these factors contribute to trends in hailstorm occurrence.

## Methods

### Data

We use three hail datasets in this study: (1) Ancient hail damage records (886 BC–1948 CE; Fig. 1a, b). These data are documented in the traditional Chinese lunisolar calendar, providing information on hail events with year and month details, while only a few entries specify the exact day. The Chinese lunisolar calendar is based on both lunar and solar motions and is designed to align seasonal variations (governed by the solar cycle) with the lunar year. Each month consists of either 29 or 30 days, resulting in a year length of ~354 or 355 days—about 11 days shorter than a solar year. To compensate for this discrepancy, a leap month is introduced into the Chinese lunisolar calendar approximately every two to three years. Over the long term, this adjustment causes the Chinese months to lag behind their corresponding Gregorian months by about one month on average<sup>50–52</sup>. Moreover, the ancient hail records have been relocated to current administrative province, ensuring consistency in spatial resolution with the current records. (2) Current hail damage records (1949–2000; grey shading in Fig. 1c), collected by the Civil Affairs Department of China. These records are based on the solar calendar. (3) Continuous and coherent hailstorm records from more than 2000 meteorological stations (1950–2010; coloured solid lines in Fig. 1c). This dataset is provided by the Chinese Meteorological Administration<sup>9,13,15</sup> and based on the solar calendar. The meteorological stations are located in each administrative county shown in Fig. S1. In addition to the occurrence of hailstorms, the maximum size of hailstones was also regularly recorded in meteorological station reports after 1980<sup>12</sup> (Fig. 1c). Hailstone size only occasionally appeared in ancient and current hail damage reports; thus, only hail size data from meteorological station reports after 1980 were used in this study.

Population datasets used to remove potential artificial effect include: (1) Historical population data (pre-1949), collected from a previous study<sup>33</sup>, with irregular time intervals. (2) Current population data (post-1949) with yearly interval, obtained from the National Bureau of Statistics of China (<https://data.stats.gov.cn>).

For temperature before 1949, five historical datasets from the National Climatic Data Center (NCDC)<sup>24–28</sup> are used. The spatial and temporal coverages of the data are listed in Table S1. The annual temperature time series is shown in Fig. S2. Among the five datasets, Mann's data cover the globe, whereas the other four cover the Northern Hemisphere. For the current period (1949–2000), temperature data are obtained from NASA/GISS (<https://climate.nasa.gov>).

Three PDO datasets are used for different purposes. To analyse the influence of PDO on the variabilities of ancient hailstorm days, we use a reconstructed annual PDO index (1470–2000), which has been validated against other reconstructed PDO indices and observational data, confirming its reliability<sup>41,54–58</sup>. For the current period (1950–2013), the PDO index is obtained from HadISST 1.1. Future PDO projection is directly sourced from the CMIP5 composite model output provided in a previous study<sup>47</sup>. This future PDO index is constructed based on the statistical synthesis of selected CMIP5 simulations and validated using historical observations to ensure consistency in phase transitions and early warning signals.

NAO data used in this paper from 1500–1949 is from the NCDC, contributed by Ortega et al.<sup>59</sup>.

### Definition of hailstorm days

A hailstorm damage day is defined as a day in which there is at least one hail damage record (ancient or current damage) in a province. A

hailstorm day in the current climate is defined as a day in which at least one storm is recorded in a province. It is important to note that a single hailstorm affecting multiple provinces will be counted multiple times. However, given China's vast geographical extent, hailfall in different provinces on the same day can also be caused by different hailstorm systems. Considering such cases, we used this method to ensure consistency and reliability in defining hail days within our dataset.

### Data agreement analysis

The data agreement was tested first before analysis. After 1949, far more hailstorm days were noted in the meteorological station reports than from current hail damage reports (Fig. 1c). However, hail damage is recorded only when damage occurs, which is largely related to the size of a hailstone. The interannual variations in the number of hailstorm days from current hail damage reports are compared with those with different bin size distributions from meteorological station records. Number variation with hailstone sizes equal to or greater than 7 mm are found to agree well with that of records from damage reports after 1980 (Fig. 1c), and the correlation coefficient is 0.85. As the damage caused by hailstones is related to their kinetic energies when they reach ground, which is a function of the size of the hailstone and can be estimated as

$$KE = 0.0243D^{3.55} \quad (1)$$

where  $D$  is the diameter of the hailstone and  $KE$  is the kinetic energy<sup>60</sup>.

The numbers of hailstorm (damage) days from the three datasets all show significant seasonal variations, with one peak in the warm season (Fig. S3). Both the current hail damage record (1949–2000) and CMA (China Meteorological Administration) records show a peak in June. However, in the ancient hail damage record, the peak appears in May, approximately one month earlier than in the current records. Since the ancient hail damage records were based on the lunisolar calendar, which is ~1 month shorter than the solar calendar currently used, this one-month shift in peak timing is expected. The lunisolar months generally lag behind their corresponding solar months by about one month. The overall agreement in seasonal variation between the hail damage records and CMA observations suggests that further valid comparisons between the current and historical hailstorm trends can be made. It is worth noting that different record location may also play a role in seasonal cycle differences since different regions have different seasonal patterns. However, its influences in our case is relatively limited, as the ancient hail damage data has already been relocated to the current administrative province.

### Detrend method

The linear regression method was used to estimate the possible influence of the population on hail damage days and remove the signals associated with artificial impacts<sup>21,22</sup>. The linear relation can be estimated as:

$$H(t) = H_0 + \beta P(t) + R(t) \quad (2)$$

where  $H(t)$  is the number of hailstorm damage days;  $H_0$  is a constant;  $P(t)$  is the population; and  $R(t)$  is the residual number of hailstorm days.

### Decomposition method

The complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method can be used to decompose the time series of a variable into various time scale components. This algorithm is a continuation of the empirical modal decomposition (EMD)<sup>29</sup> and empirical mode decomposition (EMD)<sup>61</sup> algorithms. It can overcome the problem of “mode mixing” and provide a more stable and exact reconstruction of the original signal at a lower computational cost than

those of the individual methods. In this study, the standard deviation of the Gaussian random numbers used as additional noise was set to 0.2. The number of realizations was 500 and maximum number of sifting iterations allowed was 5000. The detrend hailstorm day records, temperature anomalies, PDO and NAO records were decomposed into several intrinsic mode functions (IMFs), and the results are shown in Tables S2, S3 and S4, respectively. According to the mean periods of these modes, we obtained five major components: the interannual variation (mean period less than 10 yr, summing IMF1, IMF2, and IMF3), the interdecadal variation (mean period between 10 and 50 yr, summing IMF4 and IMF5), the multidecadal variation (MDV; IMF6), the quasicentennial variation (QCV; IMF7), and the nonlinear trend. The five major sets of time scale decomposition results are shown in Figs. S5, 6 and 7.

### Future projections method

The CEEMDAN-CNN-BiLSTM model was applied to conduct future projections of the number of hailstorm days. The CEEMDAN method was first used to decompose hailstorm days, temperatures, and PDO records into different mode components. Then, the trend component and oscillation (oscillation in this section refers to the summation of the MDV and QCV) component of hailstorms were separately predicted with a one-dimensional (1D) convolutional neural network (CNN) machine learning method<sup>43,44,46,62</sup> and a 1D CNN-based bidirectional long short-term memory (BiLSTM) parallel model with an attention mechanism<sup>63</sup>. The nonlinear trend of the global temperature anomaly was chosen as the training feature for the prediction of the hailstorm trend, whereas the PDO oscillation was used as the training feature for the prediction of the oscillation in the number of hailstorm days. Compared with 2D CNNs, 1D CNNs are preferable for processing 1D signals<sup>46</sup>. Furthermore, to better capture the hidden features of the oscillations and produce better predictions, the coupled CNN-BiLSTM-Attention model<sup>63</sup> was applied. These methods were chosen on the basis of trial and error tests.

Data before 1949 were used to train the machine learning model. The data were separated randomly into training and test sets at a ratio of 8:2. For the 1D CNN model used in trend prediction, we designed a convolutional layer with 32 filter maps and a kernel size of 1, followed by a flattening layer and a fully connected dense layer. The model was compiled using the Adam optimizer with a learning rate of 0.001 and initial decay rates  $\beta_1$  and  $\beta_2$  set to 0.9 and 0.999. The loss function used was mean squared error loss. The model was trained for up to 1000 epochs with a batch size of 16. To prevent overfitting and improve computational efficiency, early stopping was implemented. The training will be terminated if the validation loss showed no improvement after 20 consecutive epochs. We also employed five-fold cross-validation to evaluate model robustness. The best-performing model from each fold was saved to ensure that only the most optimal model was retained.

For the CNN-BiLSTM-Attention model used in oscillation prediction, we first designed a 1D convolutional layer with 32 filters and a kernel size of 1, similar to the structure in the trend prediction model. Then, a BiLSTM layer which consists of 32 hidden units was used to process the extracted features from the convolution layer, capturing temporal dependencies in both forward and backward directions. To avoid overfitting, a dropout layer with a rate of 0.5 was applied after the BiLSTM layer. An attention mechanism is then applied to dynamically weight important time steps, enhancing interpretability. Finally, a fully connected dense layer with a linear activation function was used for regression. The fitting and optimization methods for this coupled model were the same as those used for the 1D CNN model, except that the batch size was set to 8.

To eliminate the random error that may be caused by randomly splitting the data, the splitting and training processes were repeated 200 times to generate 200 ensemble models for trend and/or oscillation prediction. The ensemble mean *R* square values were 0.47/0.50

for the trend prediction and 0.78/0.76 for the oscillation prediction for training/testing.

Data between 1950 and 2099 were used for projection, and we assumed that the influence of population growth remained constant at the mean value between 1950 and 1980. The temperature anomaly between 1950 and 2000 was obtained from NASA/GISS data, and the data from 2000 to 2100 were from the ensemble mean model output of CMIP5. The temperature data were from CMIP5 runs with four different scenarios used (<https://esgf-node.llnl.gov/search/cmip5/>). The temperature anomaly and PDO index between 1950 and 2000 were first processed with CEEMDAN, and the corresponding trend and oscillation components were used for projection. The trend and oscillation predictions were subsequently summed, and the results are shown in Fig. 4. The prediction results from 1950–2000 were in good agreement with the damage records, suggesting that the CNN-BiLSTM model is capable of accurately predicting the number of future hailstorm days in China.

### Data availability

All processed data presented in the figures of this study are publicly available at GitHub: <https://github.com/atrmli/data> and have been archived on Zenodo at <https://doi.org/10.5281/zenodo.15871682><sup>64</sup>. The raw historical hailstorm records are available upon request from the corresponding author. The station-based hailstorm records (1950–2010) are available from the China Meteorological Administration (CMA) upon request ([datacenter@cma.gov.cn](mailto:datacenter@cma.gov.cn)).

### Code availability

The necessary codes for generating the figures in this study are publicly available under CC BY-NC 4.0 license at GitHub: <https://github.com/atrmli/codes>. The version of the code associated with this manuscript has been archived on Zenodo at <https://doi.org/10.5281/zenodo.15871680><sup>65</sup>.

### References

- Allen, J. T. et al. Understanding hail in the Earth system. *Rev. Geophys.* **58**, e2019RG000665 (2020).
- Raupach, T. H. et al. The effects of climate change on hailstorms. *Nat. Rev. Earth Environ.* **2**, 213–226 (2021).
- Stocker, T. *Climate Change 2013: The Physical Science Basis*. <https://www.ipcc.ch/report/ar5/wg1/> (2014).
- IPCC. *Climate Change 2021: The Physical Science Basis*. <https://www.ipcc.ch/report/ar6/wg1/> (2021).
- IPCC. *Climate Change 2014: Synthesis Report* [https://www.ipcc.ch/site/assets/uploads/2018/05/SYR\\_AR5\\_FINAL\\_full\\_wcover.pdf](https://www.ipcc.ch/site/assets/uploads/2018/05/SYR_AR5_FINAL_full_wcover.pdf) (2014).
- Changnon, S. A. & Changnon, D. Long-term fluctuations in hail incidences in the United States. *J. Clim.* **13**, 658–664 (2000).
- Allen, J. T. & Tippet, M. K. The characteristics of United States hail reports: 1955–2014. *J. Sev. Storms Meteorol.* **10**, 1–31 (2015).
- Jin, H.-G., Lee, H., Lkhamjav, J. & Baik, J.-J. A hail climatology in South Korea. *Atmos. Res.* **188**, 90–99 (2017).
- Li, M., Zhang, Q. & Zhang, F. Hail day frequency trends and associated atmospheric circulation patterns over China during 1960–2012. *J. Clim.* **29**, 7027–7044 (2016).
- Li, X. et al. Climatology of hail frequency and size in China, 1980–2015. *J. Appl. Meteorol. Climatol.* **57**, 875–887 (2018).
- Lkhamjav, J., Jin, H.-G., Lee, H. & Baik, J.-J. A hail climatology in Mongolia. *Asia Pac. J. Atmos. Sci.* **53**, 501–509 (2017).
- Ni, X., Muehlbauer, A., Allen, J. T., Zhang, Q. & Fan, J. A climatology and extreme value analysis of large hail in China. *Mon. Weather Rev.* **148**, 1431–1447 (2020).
- Ni, X. et al. Decreased hail size in China since 1980. *Sci. Rep.* **7**, 1–6 (2017).
- Xie, B., Zhang, Q., Wang, Y. Trends in hail in China during 1960–2005. *Geophys. Res. Lett.* **35**, 13 (2008).

15. Zhang, Q., Ni, X. & Zhang, F. Decreasing trend in severe weather occurrence over China during the past 50 years. *Sci. Rep.* **7**, 1–8 (2017).
16. Fei, J. Meteorological history and historical climate of China. *ORE. Clim. Sci.* <https://doi.org/10.1093/acrefore/9780190228620.013.594> (2018).
17. Chen, S., Luo, Z. & Pan, X. Natural disasters in China: 1900–2011. *Nat. Hazards* **69**, 1597–1605 (2013).
18. Bai, S. *General History of China*, Vol 91st edn (Shanghai People's Publishing House, 1999).
19. Lee, H. F. & Zhang, D. D. Natural disasters in northwestern China, AD 1270–1949. *Clim. Res.* **41**, 245–257 (2010).
20. Allen, J. T., Tippett, M. K. & Sobel, A. H. An empirical model relating U.S. monthly hail occurrence to large-scale meteorological environment. *J. Adv. Model. Earth Syst.* **7**, 226–243 (2015).
21. Mann, M. E. & Emanuel, K. A. Atlantic hurricane trends linked to climate change. *Eos Trans. AGU* **87**, 233–241 (2006).
22. Ting, M., Kushnir, Y., Seager, R. & Li, C. Forced and internal twentieth-century SST trends in the North Atlantic. *J. Clim.* **22**, 1469–1481 (2009).
23. Zhou, Z., Zhang, Q., Allen, J. T., Ni, X. & Ng, C. P. How many types of severe hailstorm environments are there globally? *Geophys. Res. Lett.* **48**, e2021GL095485 (2021).
24. Feng, S. et al. Northern Hemisphere temperature reconstruction during the last millennium using multiple annual proxies. *Clim. Res.* **56**, 231–244 (2013).
25. Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M. & Karlen, W. Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature* **433**, 613–617 (2005).
26. D'Arrigo, R., Wilson, R. & Jacoby, G. On the long-term context for late twentieth century warming. *J. Geophys. Res.* **111**, D3 (2006).
27. Hegerl, G. C. et al. Detection of human influence on a new, validated 1500 year temperature reconstruction. *J. Clim.* **20**, 650–666 (2007).
28. Mann, M. E. et al. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl. Acad. Sci. USA* **105**, 13252–13257 (2008).
29. Wu, Z. & Huang, N. E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Data Sci. Adapt. Anal.* **1**, 1–41 (2009).
30. Torres, M. E., Colominas, M. A., Schlotthauer, G. & Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing* 4144–4147 (ICASSP, 2011).
31. Shen, Z.-Y. & Ban, W.-C. Machine learning model combined with CEEMDAN algorithm for monthly precipitation prediction. *Earth Sci. Inform.* **16**, 1821–1833 (2023).
32. Bo-Tao, Z. Weakening of winter North Atlantic oscillation signal in spring precipitation over Southern China. *Atmos. Ocean. Sci. Lett.* **6**, 248–252 (2013).
33. Han-Lie, X., Juan, F. & Cheng, S. Impact of preceding summer north atlantic oscillation on early autumn precipitation over central China. *Atmos. Ocean. Sci. Lett.* **6**, 417–422 (2013).
34. Qian, C. & Zhou, T. Multidecadal variability of North China aridity and its relationship to PDO during 1900–2010. *J. Clim.* **27**, 1210–1222 (2014).
35. Xiao, M., Zhang, Q. & Singh, V. P. Influences of ENSO, NAO, IOD and PDO on seasonal precipitation regimes in the Yangtze River basin, China. *Int. J. Climatol.* **35**, 3556–3567 (2015).
36. Gao, T., Wang, H. J. & Zhou, T. Changes of extreme precipitation and nonlinear influence of climate variables over monsoon region in China. *Atmos. Res.* **197**, 379–389 (2017).
37. Lou, M., Li, C., Hao, S. & Liu, J. Variations of winter precipitation over Southeastern China in association with the North Atlantic oscillation. *J. Meteorol. Res.* **31**, 476–489 (2017).
38. Peng, Y. Simulated interannual teleconnection between the summer North Atlantic Oscillation and summer precipitation in eastern China during the last millennium. *Geophys. Res. Lett.* **45**, 7741–7747 (2018).
39. Zhou, X., Jiang, D. & Lang, X. Unstable relationship between the Pacific decadal oscillation and Eastern China summer precipitation: insights from the medieval climate anomaly and little ice age. *Holocene* **30**, 799–809 (2020).
40. Xue, J. et al. The combined influences of solar radiation and PDO on precipitation over Eastern China during the last millennium. *Clim. Dyn.* **60**, 1137–1150 (2022).
41. Shen, C., Wang, W.-C., Gong, W. & Hao, Z. A Pacific Decadal Oscillation record since 1470 AD reconstructed from proxy data of summer rainfall over eastern China. *Geophys. Res. Lett.* **33**, 3 (2006).
42. Jiang, N., Yan, Q., Xu, Z., Wang, H. Divergent response of eastern China precipitation to the Pacific decadal oscillation during the last Interglacial. *Quat. Sci. Rev.* **319**, 108344 (2023).
43. Kiranyaz, S., Ince, T., Hamila, R., Gabbouj, M. Convolutional neural networks for patient-specific ECG classification. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2608–2611 (IEEE, 2015).
44. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017).
45. Qin, D. et al. A novel combined prediction scheme based on CNN and LSTM for urban PM2.5 concentration. *IEEE Access* **7**, 20050–20059 (2019).
46. Kiranyaz, S. et al. 1D convolutional neural networks and applications: a survey. *Mech. Syst. Signal Process.* **151**, 107398 (2021).
47. Hao, W., Pengcheng, Y., Wei, H., Junhu, Z. & Guolin, F. Detection of decadal phase transition and early warning signals of PDO in recent and next 100 years. *Chin. J. Atmos. Sci.* **46**, 225–236 (2022).
48. Blair, S. & Leighton, J. Creating high-resolution hail datasets using social media and post-storm ground surveys. *Electron. J. Oper. Meteorol.* **13**, 32–45 (2012).
49. Tuovinen, J.-P., Punkka, A.-J., Rauhala, J., Hohti, H. & Schultz, D. M. Climatology of severe hail in Finland: 1930–2006. *Mon. Weather Rev.* **137**, 2238–2249 (2009).
50. Olaniyi, B. A. Historical analysis of calendars-Chinese calendars and world calendars. *Asia. Jour. Res. Soci. Scie. Human.* **4**, 114–130 (2014).
51. Aslaksen, H. *The Mathematics of the Chinese Calendar*. (National University of Singapore, 2010).
52. Jiang, X. *The Studies of Heaven and Earth in Ancient China* 1st edn, Vol. 521 (Springer, 2021).
53. Deng, K. G. Unveiling China's true population statistics for the pre-modern era with official census data. *Popul. Rev.* **43**, 32–69 (2004).
54. Porter, S. E., Mosley-Thompson, E., Thompson, L. G. & Wilson, A. B. Reconstructing an interdecadal Pacific oscillation index from a Pacific basin-wide collection of ice core records. *J. Clim.* **34**, 3839–3852 (2021).
55. Grove, C. A. et al. Madagascar corals reveal a multidecadal signature of rainfall and river runoff since 1708. *Clim.* **9**, 641–656 (2013).
56. Henley, B. J. Pacific decadal climate variability: Indices, patterns and tropical-extratropical interactions. *Glob. Planet. Change* **155**, 42–55 (2017).
57. Heyerdahl, E. K., Morgan, P. & Riser, I. I. J. P. Multi-season climate synchronized historical fires in dry forests (1650–1900), northern rockies, USA. *Ecology* **89**, 705–716 (2008).
58. Landrum, L. et al. Last Millennium Climate and Its Variability in CCSM4. *J. Clim.* **26**, 1085–1111 (2013).
59. Ortega, P. et al. A model-tested North Atlantic oscillation reconstruction for the past millennium. *Nature* **523**, 71–74 (2015).

60. Heymsfield, A., Szakáll, M., Jost, A., Giammanco, I. & Wright, R. A comprehensive observational study of graupel and hail terminal velocity, mass flux, and kinetic energy. *J. Atmos. Sci.* **75**, 3861–3885 (2018).
61. Huang, N. E. et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **454**, 903–995 (1998).
62. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
63. Yin, X., Liu, Z., Liu, D. & Ren, X. A Novel CNN-based Bi-LSTM parallel model with attention mechanism for human activity recognition with noisy data. *Sci. Rep.* **12**, 7878 (2022).
64. Zhang, Q. et al. Data for “climate impacts and future trends of hailstorms in china based on millennial records”. *Zenodo* <https://doi.org/10.5281/zenodo.15871682> (2025).
65. Zhang, Q. et al. Code for “Data for “climate impacts and future trends of hailstorms in china based on millennial records”. *Zenodo* <https://doi.org/10.5281/zenodo.15871680> (2025).

## Acknowledgements

We thank Xiaomeng Huang for his help in guiding the machine learning analysis. This work was supported by the National Natural Science Foundation of China (Grant No. 42030607). We thank for the technical support of the National Large Scientific and Technological Infrastructure “Earth System Numerical Simulation Facility” (<https://cstr.cn/31134.02.EL>).

## Author contributions

Q.Z. and W.L. conceived the paper. Q.Z., R.L., X.L., C.-P.N. and C.Z. carried out the literature searches and read the papers. R.L. and X.L. analysed the data. R.L. and S.Z. created the figures. Q.Z. wrote the first draft of the paper. W. L. and R. L. helped revise the final version of the article and were responsible for the decision to submit the article. All the authors collected and verified the data. R.L., Q.Z. and W.L. confirm that they have full access to all the data in the study and accept responsibility for this publication.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-63028-7>.

**Correspondence** and requests for materials should be addressed to Qinghong Zhang or Wenhong Li.

**Peer review information** *Nature Communications* thanks Yaojie Yue and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025