ARTICLE   OPEN

Check for updates

Using particle swarm optimization to improve visibility-aerosol optical depth retrieval method

Jian Wu [1]✉, Shuang Zhang[1], Qidong Yang[1], Deming Zhao[2], Wenxuan Fan[1], Jingchuan Zhao[1] and Cheng Shen [3]

In view of the lack of long-term AOD (Aerosol Optical Depth) data, PSO (Particle Swarm Optimization) algorithm is introduced and joint used with NLSM (the nonlinear least square method) to improve visibility-AOD retrieval method, which is referred to as the PSO-M-Elterman model and significantly increases data available rate by 8% and correlation by about 20% with the true value in the experimental group. The mean absolute error, the proportion of the smaller absolute error and the root mean square error in the PSO-M-Elterman model experimental group are 0.0314 and 91.23%, 0.0509 respectively, which significantly outperforms other groups. The main increase of AOD was found in the eastern region (South China, East China, Central China) and Taklimakan with the trend coefficients of 2.67, 2.46, 2.13, and 1.45 ($\times 10^{-3}$ yr^{-1}) in recent 55 years, which may not be interpreted by the influence of relative humidity. Long-term change of AOD in east China is mainly caused by human activity, and the AOD is higher in cities with a larger population and more human activity. The PSO-M-Elterman model can maximize the advantage of visibility sequence length to obtain long-term AOD inversion results.

INTRODUCTION

Aerosols originate from a wide variety of sources, and often have highly complex compositions, which play an important role in global and regional environments. Understanding the physical, chemical, and optical characteristics and distribution of aerosols aids to assess their role in radiation balance, air quality, and predict climate change[1,2]. The composition, concentration, absorption, and scattering characteristics of the aerosol are directly or indirectly observed by remote sensing and ground-based sensors[3–7]. The aerosol optical depth (AOD), one of the most important optical parameters of aerosols, is a key physical quantity characterizing atmospheric turbidity, and an important factor used to determine the effect of aerosols on the climate. The space coverage of satellite data is wide; however, the timelapse is short, whereas ground-based data acquired by a sunphotometer provide high accuracy, but few stations. Neither can provide AOD data spanning a wide range, long time, and high consistency[8].

Currently, studies on the long-term change of AOD are scarce due to limited aerosol data. Several studies revealed an interaction between aerosols and monsoon systems[9–12]. Therefore, obtaining long-term aerosol data to analyze the interaction between aerosols and climate change factors is an urgent issue in the effort to understand the impact of human activities on the climate. AOD can be calculated from the extinction coefficient, wide-band solar radiation, visibility, and other meteorological elements[13–16]. Visibility, as an indicator reflecting atmospheric transparency, has the advantage of long-term observation and numerous stations, and it is often used to retrieve AOD and analyze its spatial and temporal characteristics. AOD retrieval using visibility data shows that the Indian Peninsula, Central Europe, West Africa, Southeast North America, and Southeast China are the largest AOD regions in the world[17]. The spatial distribution of inversion AOD by visibility in China is high in the southeast and low in the northwest, while the seasonal distribution is high in summer and low in winter, showing an overall increasing trend in recent decades[18–21].

AOD retrieval methods worldwide mostly use the original visibility- AOD conversion model[17,22,23]. The inversion method is based on the Elterman model and takes into account the influence of water vapor. A set of empirical parameters are introduced into each station and developed into the M-Elterman and the KM-Elterman models[20,21,24], whose retrieval parameters are determined by the nonlinear least square method (NLSM)[20]. These AOD retrieval methods using NLSM suffer from inaccurate parameters, which limit their wider application. Further, Lin[25,26] proposed an AOD retrieval method using chemical transmission model GEOS-Chem fusion visibility data, and the results were in good agreement with MODIS data.

To address the inaccurate parameters in the M-Elterman model, this study introduces the particle swarm optimization (PSO) algorithm into the AOD retrieval method, which originates from a study on the foraging behavior of birds[27]. Owing to the simplicity, stability, and efficiency of this algorithm, as well as its ability to obtain global or approximately global optimal solutions of the problem[28], it has been widely applied across different fields, and its use has become particularly widespread in the field of artificial intelligence[29–31]. The essence of the PSO algorithm is that one or more functions (usually nonlinear) can solve the problem of the global optimal solution in a multi-dimensional space. The most significant difference between this and the traditional algorithm, which uses tools such as derivatives to find extrema, is that the initial value, the update of particle velocity and position are generated randomly. Under the constraint of an objective function, as long as there is a sufficient number of iterations, these particles will tend to converge, and thus obtain the optimal solution of the problem.

The PSO algorithm is implemented by initializing the parameters to be optimized to the position of a group of random

[1]Key Laboratory of Atmospheric Environment and Processes in the Boundary Layer over the Low-Latitude Plateau Region, Department of Atmospheric Science, Yunnan University, Kunming 650091, China. [2]CAS Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China. [3]Gaochun Meteorological Bureau, Nanjing 211300, China. ✉email: wujian@ynu.edu.cn
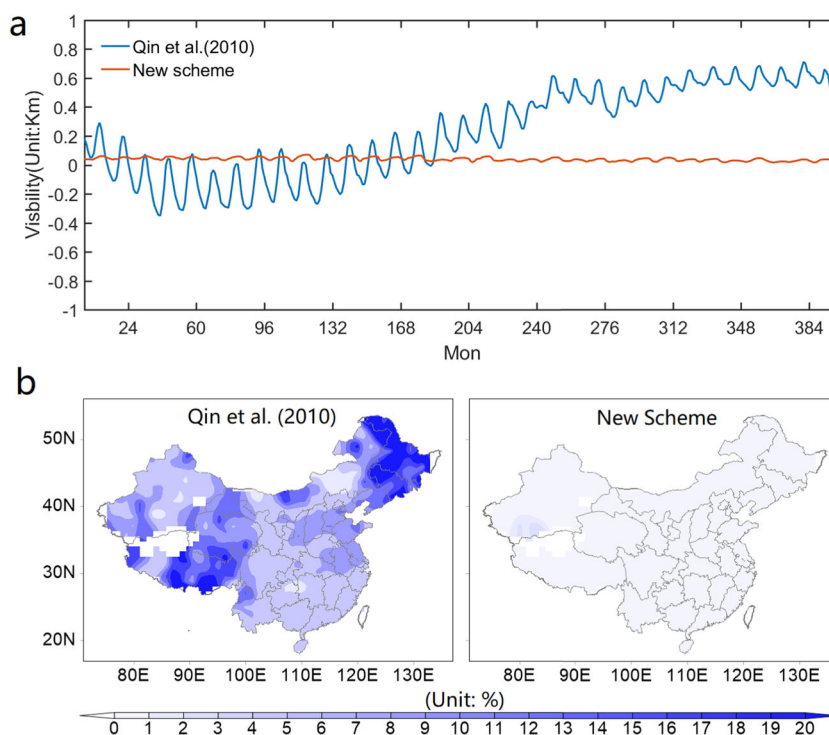
Fig. 1 Comparison of four schemes for visibility grade conversion. a Monthly time series of national mean visibility error from 1980–2013. **b** Spatial distribution of mean visibility relative error from 1980 to 2013.

particles, and subsequently finding the optimal solution through iteration. In each step of the iteration, the particle updates its position by tracking two extrema, one being the optimal solution found by the particle itself, which is called the individual extrema, while the other is the optimal solution found by the entire population, which is the global extrema. After finding these two extrema, the particle repeatedly updates its velocity and position, eventually reaching the global extreme for all particles. Compared with the NLSM, the optimization of parameters is considered to be a similar process, as random parameter combinations are selected to minimize the deviation between the observed and simulated values within a certain period of time, and the parameter combination with the smallest deviation is considered as the optimal parameter. Both algorithms require iterative calculation; the difference lies in the randomized initial parameters of the PSO algorithm. In the actual application, because the calculation time is controlled (small storage requirements), the value range of the parameter is given, and the parameter space to be optimized becomes a closed area. Thus, the global optimal parameters in this area can be obtained, which represent the approximate optimal parameters in the entire open area.

RESULTS

Conversion of visibility grade data

Visibility was recorded on a scale of ten before 1980, and the actual visibility distance (unit: km) was directly recorded after 1980. Therefore, it is necessary to convert the visibility grade data before 1980 into the corresponding distance data. We designed a conversion scheme based on the scheme of Qin[19] (details are in the Methods section), which is commonly used in visibility conversion, and compared our results. To estimate alternative distances in different schemes, we converted the visibility distance data from 1980 to 2013 into their corresponding grade, and subsequently converted them into distances

according to the conversion schemes described above. Finally, we compared them with the actual observed visibility distance, whose difference is considered to be the error between the converted visibility using the conversion scheme and the actual visibility, and the relative error is the absolute error divided by the actual visibility. The comparison of the mean error in the monthly series (Fig. 1a) shows that the errors fluctuated periodically over a period of 12 months. The absolute error of scheme Qin fluctuates within the range of −0.4–1 km. In contrast, the mean absolute error of the conversion proposed scheme only fluctuates around zero with an amplitude below 0.1 km, and does not change significantly with time. Furthermore, the distribution of the relative error (Fig. 1b) shows that the Qin scheme exhibits the highest relative error in terms of visibility, generally above 10%, whereas in the conversion scheme, except for a few areas in the southwest of Xinjiang, the relative error is below 1%.

Hence, we demonstrate that the conversion scheme significantly improves the credibility of visibility conversion owing to the consideration of differences and seasonal characteristics of each station. The improved converted visibility provides more reliable data for the subsequent AOD inversion, and further improves the accuracy of the inversion results. This scheme can also be applied to other scientific research that requires long-term visibility data in China, as it can effectively improve the homogeneity of visibility.

Joint use of inversion parameters from different algorithms

To verify the reliability and accuracy of the inversion, we split the period of satellite observation AOD into an experimental and an extrapolation verification group, containing data collected over 94 months from January 2007 to October 2014 and over 72 months from January 2001 to December 2006, respectively. For each station, the parameters were first fitted using the PSO algorithm and then filtered through the combined criterion
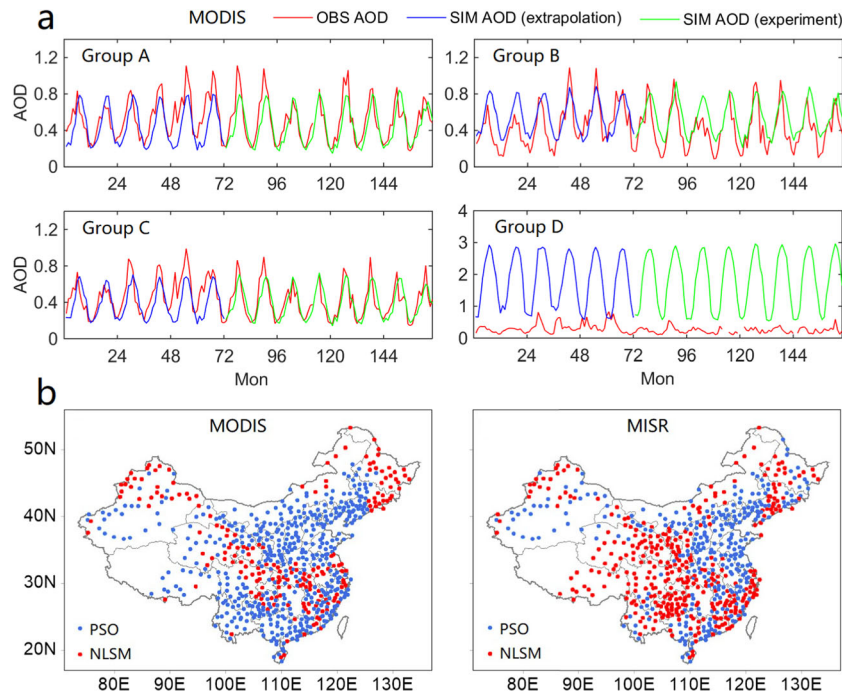
Fig. 2 Selection of parameter simulation algorithm. a Time series of inversion AOD and true value in different evaluation intervals when using MODIS data. **b** Distribution of station selection for PSO and NLSM algorithm in PSO-M-Elterman model.

of KGE (Kling-Gupta efficiency)[32] and R:

$$KGE = 1 - \sqrt{(R-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \qquad (1)$$

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \times \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (2)$$

$$\alpha = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}}{\sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}}} \qquad (3)$$

$$\beta = \frac{\bar{x}}{\bar{y}} \qquad (4)$$

where x and y denote the true and inverted AOD, respectively.

According to the definitions of KGE and R, their larger values contribute to a better fitting effect. To evaluate retrieval results, we divided the values of KGE and R of each site into four different evaluation intervals- KGE & $R > 0.7$; KGE & $R > 0.5$; KGE & $R > 0.3$ and KGE & $R > 0$ (Corresponding to "excellent -A", "good -B", "medium -C", "poor -D"), then evaluate inversion values in each interval respectively (When KGE or $R < 0$, the inversion results can be considered to be completely unavailable).

Figure 2a shows the comparison of the time series of inversion AOD and the true value of four evaluation intervals when MODIS data are used as the AOD truth value. The values and fluctuation characteristics of the inversion AOD are similar to the true value in the first three KGE and R evaluation intervals, while the inversion AOD is significantly higher than the true value in the fourth evaluation interval with poor temporal variations. Therefore, the reliability and accuracy of inversion results are acceptable when KGE & $R > 0.3$ (i.e., the inversion results are evaluated as "medium-C" or above). Thus, the stations of KGE & $R > 0.3$ were selected to use the parameters fitted by the PSO algorithm, while other stations used the parameters fitted by NLSM in the retrieval. This joint use of retrieval parameters from PSO and NLSM in the

M-Elterman model is referred to as the PSO-M-Elterman model. In this model, 467 stations employed the PSO algorithm to determine retrieval parameters, and 194 stations employed the NLSM when MODIS data were used as the true value. Further, the PSO algorithm and NLSM were used in 308 and 353 stations respectively, when MISR data were used as true value (as shown in Fig. 2b, the selection method for specific stations is described in the discussion section of this article).

Spatiotemporal contrast of different inversion results

Figure 3 shows the mean spatial distribution and temporal fluctuation of AOD in the experimental and the extrapolation groups with MODIS and MISR data denoting the true values, respectively. The spatial distribution of MODIS/MISR AOD shows a similar distribution pattern in both the experimental and the extrapolation groups (Fig. 3a, c): southern Xinjiang, and east China (North China Plain, the middle and lower reaches of the Yangtze River, and Sichuan Basin) are the regions with high AOD values, while parts of Southwest China and Northeast China are the regions with low AOD values. However, MODIS AOD is generally higher than MISR AOD, except for some areas in the southwest (the intersecting area of Tibet, Sichuan, and Yunnan). The time series (Fig. 3b, d) show evident fluctuation characteristics with a period of one year. The AOD is higher in the summer and autumn and lowers in the spring and winter. The monthly average AOD of MODIS varies from 0.25 to 0.65 with a mean value of 0.4472, while the monthly average AOD of MISR varies from 0.15–0.5 with a mean value of 0.2928.

When MODIS data are employed as the true value (Fig. 3a), both models perform better to estimate the spatial distribution pattern of AOD. However, the AOD simulated by the M-Elterman model in Southwest Xinjiang and Inner Mongolia is approximately 0.1 higher than the true value. Meanwhile, the AOD simulated by the PSO-M-Elterman model in the boundary area of Qinghai and Xinjiang and some areas of the North China Plain is higher than 0.8. The true value of the experimental group in the southwest and northern regions (Inner Mongolia, Heilongjiang, etc.) is less than 0.1, which is lower than that of the extrapolation group.
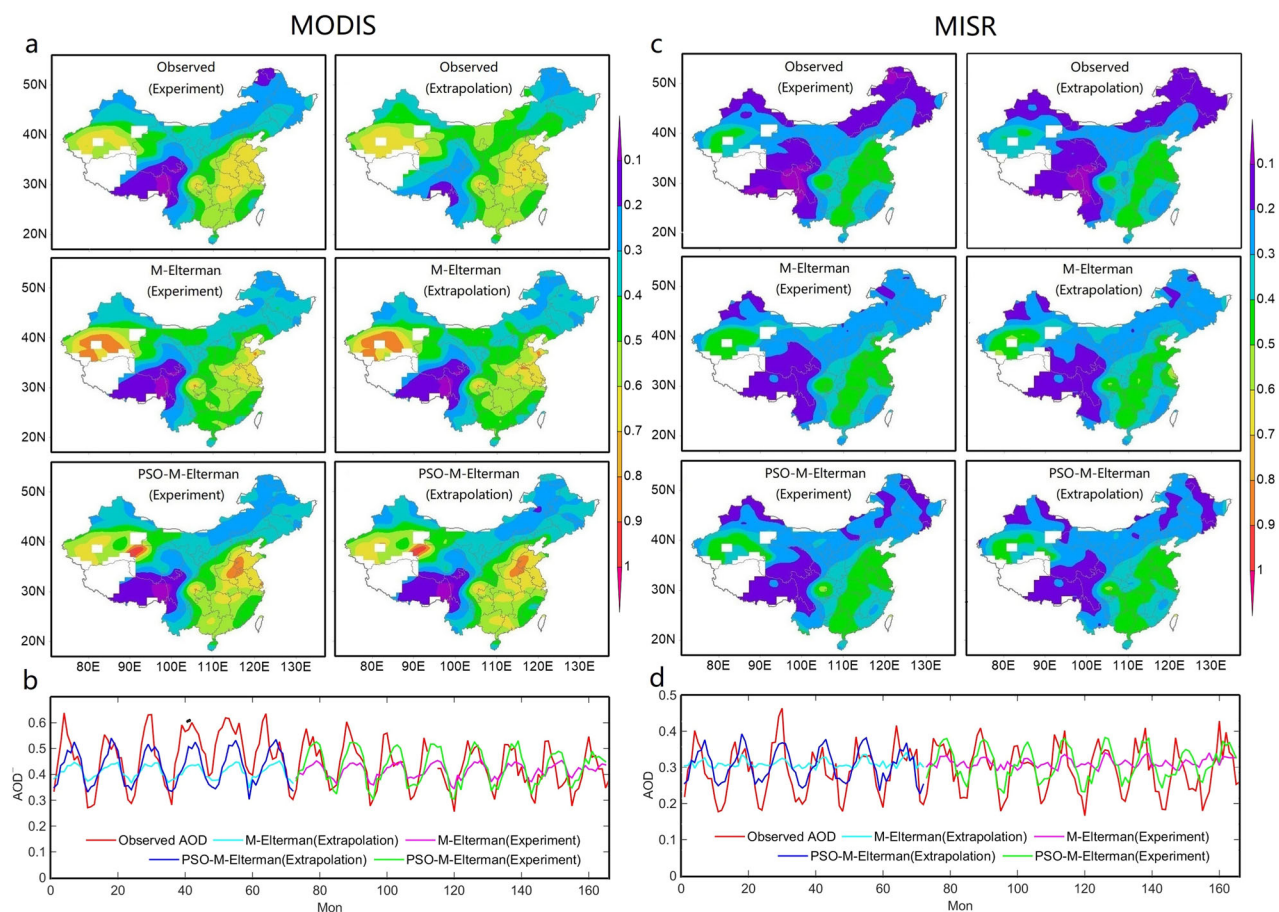
Fig. 3 **Evaluation of the mean value of inversion AOD.** The spatial distribution of observed AOD, M-Elterman AOD, PSO-M-Elterman AOD (**a**) and the monthly time series from January 2001 to October 2014 (**b**) when using MODIS data. The spatial distribution of observed AOD, M-Elterman AOD, PSO-M-Elterman AOD (**c**) and the monthly time series from January 2001 to October 2014 (**d**) when using MISR data.

For MISR data (Fig. 3c), the true value of the experimental group is slightly higher than that of the extrapolation group in Southwest Xinjiang (<0.05). The spatial distribution of AOD simulated by the M-Elterman and the PSO-M-Elterman models is similar, except for some areas in Northeast China, and the overall value is below 0.5. The retrieved AOD of the two models in northern China (Inner Mongolia, Heilongjiang) and southwestern China (the junction of Yunnan, Sichuan, and Tibet) is slightly higher than the true value. The PSO-M-Elterman model evidently captures more low-value areas of AOD, whereas areas with a true values below 0.1 are not well presented.

The performance of spatial pattern of AOD obtained by the PSO-M-Elterman model is not significantly better than that of the M-Elterman model. The spatial distribution of AOD simulated by the two models is similar, irrespective of whether MODIS or MISR data are used as AOD truth values, while there are slight differences in several regions. The performance of the extrapolation group was slightly exacerbated compared to that of the experimental group. For the time series (Fig. 3b, d), true value presents a one-year fluctuation period that first increases and then decrease (MODIS AOD fluctuate between 0.2–0.7, MISR AOD fluctuate between 0.15–0.5), which indicates that AOD is higher in summer and lower in spring and winter. But the fluctuation range of AOD inversed by the PSO-M-Elterman model is about 0.2 larger than that of the M-Elterman model, which are more consistent with the actual situation. The performance of the experimental group is likewise slightly better than that of the extrapolation group, although extreme values are not well presented in the two groups.

The spatial distribution and probability density of the correlation coefficients between the inversion and true value time series at each station are shown in Fig. 4. When using MODIS data, the positive correlation coefficients of the M-Elterman model are mainly distributed in Southwest China and North China, and most of the values are below 0.5. Almost half of the correlation coefficients in the region assume negative values. The PSO-M-Elterman model not only increases the coverage area of the positive correlation coefficient, but also enhances the correlation values. The probability peak of the correlation coefficient appears between 0.4 and 0.5, and is associated with few negative correlations. The pattern of the extrapolation group is similar, but slightly worse than that of the experimental group. The proportion of $R > 0.3$ stations in the PSO-M-Elterman experimental and extrapolation groups are 67.17% and 56.58%, respectively, which is considerably higher than that in the M-Elterman model. For MISR data, the M-Elterman model likewise yields poor correlation, and the PSO-M-Elterman model presents a slightly weaker positive correlation pattern than MODIS data. Furthermore, the area of positive correlation in the experimental group almost covers the entire region of China. The inversion results of the extrapolation group are slightly worse than that of the experimental group, and the proportion of the $R > 0.3$ stations in the two groups reach 54.16% and 47.05%, respectively, which is about 26% and 16% higher than that of the M-Elterman model, respectively. The inversion results of the PSO-M-Elterman model show significant improvement in terms of the fluctuation trend and correlation of time series relative to the M-Elterman model.
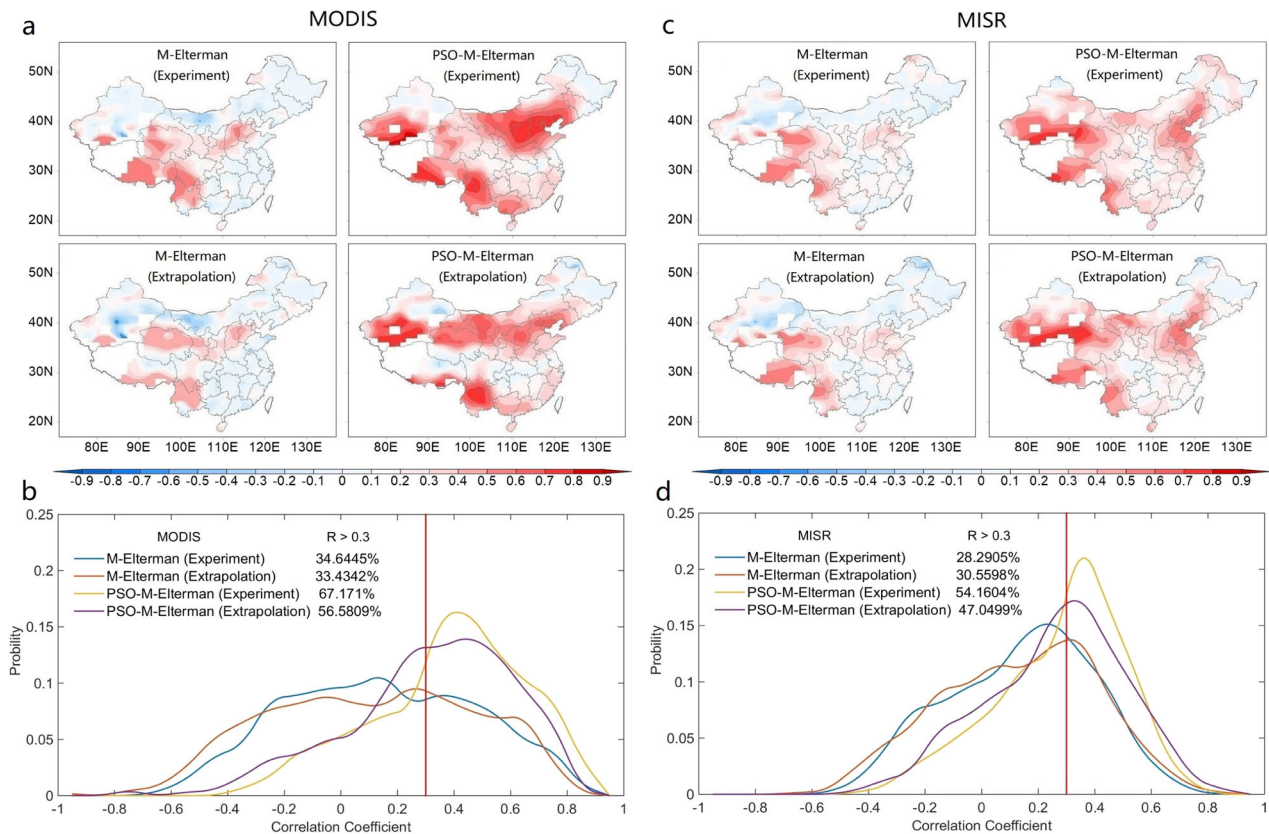
Fig. 4 Evaluation of the correlation of inversion AOD. The spatial distribution (**a**) and probability density plot (**b**) of correlation coefficients between inversion AOD in M-Elterman/PSO-M-Elterman models and observed AOD when using MODIS data. The spatial distribution (**c**) and probability density plot (**d**) of correlation coefficients between inversion AOD in M-Elterman/PSO-M-Elterman models and observed AOD when using MISR data.

Quantitative evaluation of different inversion results

Several indices (Table 1) were used to evaluate the performances of different retrieval models using different satellite data. The PSO-M-Elterman model evidently outperforms the M-Elterman model, irrespective of whether MODIS or MISR data are used as the true value. For the data unavailability rate, the M-Elterman model achieves more than 20% using MODIS data, and approximately 15% using MISR data. The PSO-M-Elterman model significantly increases data availability by 8% and 6%, respectively. The mean absolute error and the proportion of the smaller absolute error in the PSO-M-Elterman model experimental group are 0.0314 and 91.23%, respectively, and the root means square error was only 0.0509, which significantly outperforms other groups. While in the M-Elterman model, the absolute error of each group is generally greater than 0.04 and the proportion of the smaller absolute error is less than 85%. At the same time, the correlation coefficient of each group is lower than 0.4, and there are only less than 40% of sites with a correlation coefficient above 0.3. Unlike other parameters, for which the experimental group is superior to the extrapolation group, the correlation coefficients of the two models showed better results in the latter. The correlation coefficient of the PSO-M-Elterman model is the highest with MISR data, while the proportion of the stronger correlation coefficient is slightly larger with MODIS data.

Considering all evaluation indices, the PSO-M-Elterman model evidently improves the inversion results relative to the M-Elterman model with regard to all aspects, and the inversion results using MISR data outperform those using MODIS data. Further, the performance of the extrapolation group is slightly lower than that of the experimental group. The above evaluation of inversion results from different perspectives shows that the M-Elterman

algorithm simulates the average AOD state more effectively. Previous studies[20,21] focused more on the spatial distribution and correlation of the whole region average in the discussion of inversion results. However, in a smaller area or even a single station, it is difficult to depict the temporal fluctuation characteristics of AOD. Nevertheless, the PSO-M-Elterman algorithm significantly improved the inversion results, especially the regions with a small absolute error and high correlation, which indicates that it is more applicable to the visibility-AOD inversion formula than the NLSM.

Long-term changes of retrieval AOD in China

MISR AOD data were used to represent the true value, and the parameters simulated by the PSO-M-Elterman model from 2007 to 2014 were applied to the AOD retrieval before 2000. Thus, the long-term AOD series from 1960 could be established by inversion AOD. The spatial pattern and temporal series of inversion AOD from 1960 to 2014, and the spatial and temporal fields of the first two modes of EOF (Empirical Orthogonal Function) analysis are shown in Fig. 5. The general characteristics of the AOD pattern are similar to that after 2000: the high-value areas are distributed in southern Xinjiang, the Sichuan Basin, and southeastern China, while the low-value areas include northern Xinjiang, Southwest China, and Northeast China. Compared with the spatial distribution after 2000 (Fig. 3c), the high-value region (AOD > 0.5) narrowed in Southeast China, while the low-value region (AOD < 0.2) broadened slightly in northern Xinjiang (Fig. 5a).

The AOD of the whole region average gradually increased in time, which was associated with the annual fluctuation exhibiting a growth rate of 0.0011 yr^{-1}. The fluctuation range of annual AOD

Table 1. Comparison of evaluation parameters of inversion results at different satellite data/experimental groups.

Evaluation Parameters	MODIS				MISR			
	M-Elterman		PSO-M-Elterman		M-Elterman		PSO-M-Elterman	
	Experimental group	Extrapolation group	Experimental group	Extrapolation group	Experimental group	Extrapolation group	Experimental group	Extrapolation group
Data Unavailable Rate	22.44%	22.62%	8.09%	8.76%	14.76%	15.04%	6.04%	6.17%
Mean Absolute Error	0.0735	0.0985	0.0440	0.0881	0.0468	0.0634	0.0314	0.0532
Proportion of smaller absolute error (<0.1) stations	72.16%	60.67%	85.48%	63.99%	83.36%	75.34%	91.23%	83.21%
Root-Mean-Square Error	0.1140	0.1294	0.0839	0.1147	0.0668	0.0881	0.0509	0.0713
Correlation Coefficient	0.3224	0.4407	0.4369	0.4726	0.1806	0.3141	0.4824	0.5621
Proportion of stronger correlation coefficient (>0.3) stations	34.65%	33.43%	67.17%	56.08%	28.29%	30.56%	54.16%	47.05%

variation gradually increased with time (Fig. 5b). The variance contribution rates of the first two modes of EOF decomposition in AOD anomalies are 47.33% and 12.21%, respectively, (both passed the North test). The eigenvector of the first spatial mode alternates between positive and negative, which represents two trending characteristics of the AOD. In addition to the intersecting region of Sichuan, Qinghai, Tibet, southern Xinjiang, and part of Inner Mongolia, other regions are basically positive anomalies, especially in North China, the middle and lower reaches of the Yangtze River, northwestern Xinjiang, Guangxi, and Guangdong. Combined with the time coefficient of the first mode, the AOD in these areas shows an increasing trend after 1960, and the fluctuations tend to flatten after about 2005, which indicates that the anthropogenic aerosol content increased rapidly with the rapid development of the economy and population in eastern China[33]. Southern Xinjiang, Qinghai, and other regions with significant negative abnormal AOD showed a weak declining trend in the past 50 years. The second spatial distribution mode also shows the alternation between positive and negative anomalies with small values, with no significant anomaly center. Only the southwest region of the Xinjiang positive anomaly is slightly stronger, with the time coefficient first increasing and subsequently decreasing, indicating that the AOD in Southwest Xinjiang was higher from 1984 to 2000, and lower in other years.

We employed power spectrum analysis, wavelet analysis, and EEMD (Ensemble Empirical Mode Decomposition) analysis to explore the periodic characteristics of AOD in China in recent decades; however, none passed the confidence test. This indicates that the increasing trend of aerosol in China may be predominantly attributed to the local emission of polluting gases and particulate matter, rather than the long-distance transport of atmospheric circulation.

Distribution characteristics of inversion AOD in different regions

According to the average distribution of inversion AOD and EOF decomposition, China is divided into 9 regions (each region has similar distribution and trend characteristics) to further discuss the regional long-term variation characteristics of AOD. In addition, many researches pointed that as the relative humidity increases, the hygroscopic growth of aerosols will continue and become faster[34–37]. Considering that water vapor is a major factor affecting physicochemical and optical properties of aerosols[38–40], the variation characteristics of relative humidity are also discussed. (Please refer to Supplementary Fig. 1 for details of regional division).

In Fig. 6a, the regional average AOD in the Sichuan Basin reaches the highest (0.4436) in China, followed by central China (0.404). AOD fluctuates around 0.3 in South China, East China, and Taklimakan region. The lowest average AOD in northeast China and Northwest China are 0.234 and 0.235, respectively. Weak upward trends present in all regions, in which South China, East China, Central China, and the Taklimakan region passed the significance t-test at 95% confidence level with the linear trend coefficients of 2.67, 2.46, 2.13, and 1.45 ($\times 10^{-3}$ yr^{-1}) respectively. These features generally confirm the conclusions of previous research[20].

As shown in Fig. 6b, the relative humidity of all regions do not show distinct rising or descending characteristics with a small fluctuation range. Mean relative humidity are higher than 70% in East China, Central China, South China, and Sichuan Basin, where the AOD are also relatively high. But the Taklimakan has the lowest relative humidity 47.19% and the relatively high AOD 0.3381. This may be due to the fact that aerosol composition varies greatly in different regions. The hygroscopic properties of atmospheric aerosol particles influence ambient particle size, density, and mass which in turn control the lifetime and removal
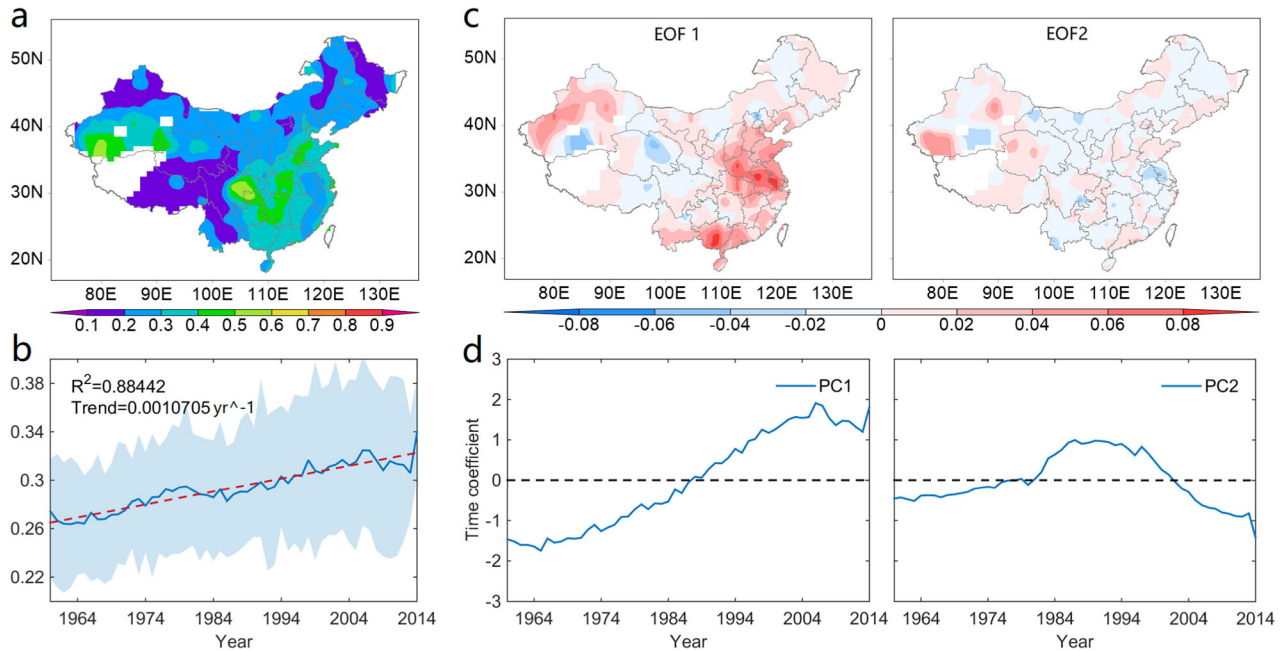
Fig. 5 Spatiotemporal characteristics of inversion AOD. a Spatial distribution of AOD during 1960–2000. **b** Annual time series of AOD during 1960–2014 (The upper and lower bounds of the shadow are the highest and lowest AOD values of the year, respectively). **c** EOF analysis of spatial field and (**d**) time coefficient series of AOD during 1960–2014.

mechanisms of the particles, and then affect the value of AOD. The higher AOD in Taklimakan is mainly caused by dust aerosols, which are hydrophobic. While the eastern region has more sulfate or nitrate aerosols caused by human emissions, which are hydrophilic and more affected by relative humidity.

The long-term variation characteristics of relative humidity and AOD in different types of large cities in east China are in Supplementary Fig. 3 for further analysis (Almost all large cities are located in the east of China, where anthropogenic emissions are higher and have more hydrophilic aerosols). The classification of city types is based on Supplementary Fig. 2a. The relative humidity in the four types of large cities all showed a weak decreasing trend. Although the downward trend is not obvious, there are strong negative correlation between relative humidity and AOD with a value of −0.6168, −0.6521, −0.6872 in ILarge City, IILarge City, and Extra Large City. This means that a decrease in relative humidity may be more conducive to aerosol accumulation. The negative correlation between AOD and relative humidity is not obvious in Super Large City (only 7 cities), which may be due to the scattered distribution of cities and different regional environmental impacts.

AOD in Sichuan Basin maintains high value all the year round (>0.4) with nearly no monthly fluctuation (Fig. 6c, d) with a maximum (83.02%) in October. There is a distinct rising phase followed by a decreasing phase for both AOD and relative humidity in North China, East China, South China, southwest China. In the Taklimakan, however, AOD is lower in winter (<0.3) and higher in spring and summer (around 0.4), which synchronizes the onset of sandstorms in spring and gradual abating in autumn, with relative humidity low in months with high AOD. Different from other regions, AOD in Northeast China reaches its peak in April (0.304) and October (0.296) respectively, and shows a fluctuation characteristic of ascending - descending - ascending - descending throughout the year, but relative humidity shows different variations with high in summer low in other seasons. The monthly characteristic of AOD and relative humidity are almost consistent in East China, North China, and Southwest China but varies in other regions, and the reasons need to be further studied.

Combined with the analysis of AOD in cities with different populations in the Supplementary Discussion, long-term change of AOD in China is mainly caused by human activity, and the AOD is higher in cities with a larger population and more human activity. Based on the analysis above, we think that the long-term increasing trend of AOD is not mainly caused by long-term changes of relative humidity, but does not deny the influence of relative humidity on aerosols, which may occur more in local Spatio-temporal scales and varies in different regions.

DISCUSSION

NLSM employs the minimum sum of error squares as the criterion to estimate the parameters of a static model. Because of the nonlinearity of the M-Elterman model, the parameter estimation cannot be obtained by calculating the extreme value of multivariate function according to the linear least square method. The M-Elterman model adopts an iterative algorithm, that is, starting from an initial value of parameters (i.e., the original parameters of the Qiu model), and subsequently generating a series of iterative parameter points. If this parameter sequence converges to a parameter point that minimizes the objective function, it will be regarded as the final solution. However, as the initial parameters are fixed, the simulated parameters are usually only optimized locally, but not applicable globally, which takes the partial inversion AOD beyond the reasonable value range. During the experiment, we found that parameters simulated in different stations are highly sensitive to variations in the PW (pressure of water vapor) value. This is because in the product term of the M-Elterman model, PW exists in two terms of the natural power (Eq. (15)), and the second term of the product term in the power is likewise an exponential relation. This means that the values of parameters b and d (especially d) determine the sensitivity of the inversion result to the PW value.

We randomly selected four different stations and used MISR data as the true value, while keeping the altitude and actual visibility unchanged to test the sensitivity of retrieval AOD to PW value between 0 and 20 hPa (Fig. 7a). The inversion AOD in stations "50968" and "50114" is within 0.4 and increases with the
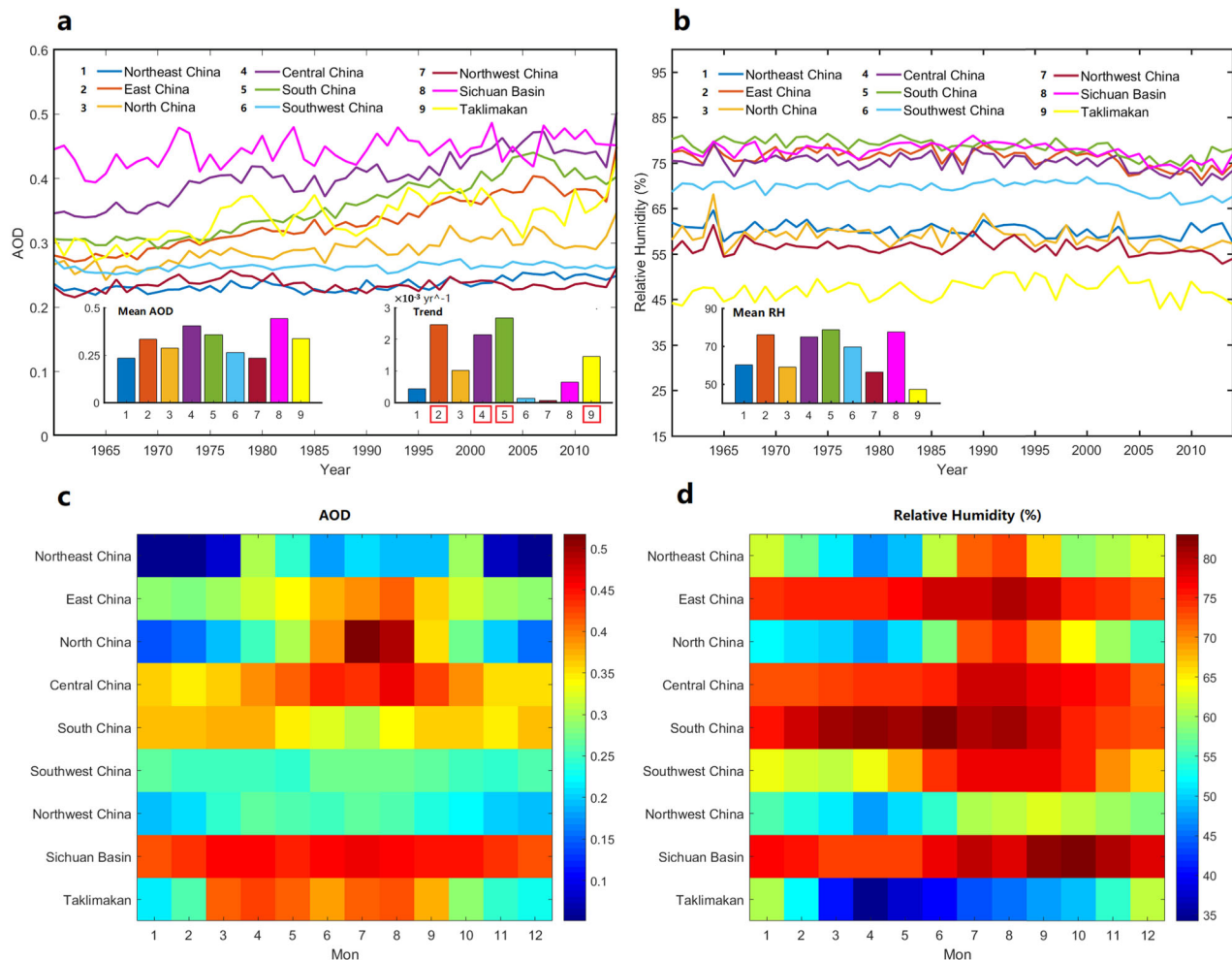
Fig. 6 Annual and monthly variations of AOD and relative humidity in different regions. a Annual series of AOD from 1960 to 2014 (The regions in the red boxes represent trend significance t-test at 95% confidence level). **b** Annual series of relative humidity from 1960 to 2014. **c** Monthly distribution of AOD. **d** Monthly distribution of relative humidity.

rise of PW, which is in line with the actual situation. However, the inversion AOD in stations "51330" and "54838" appears as an extremely large value within a normal PW range. When the PW is larger than a certain critical value, the AOD conforms to the actual situation (AOD < 1). This is because the parameters simulated by NLSM for these two stations are not applicable to the global situation, and similar situations exist in other stations and the MODIS data.

The spatial distribution of the unavailability rate of inversion AOD data using the NLSM to solve the parameters (Fig. 7b) shows there is a severe lack of AOD data in Inner Mongolia, Northeast China, and North China, and the unavailability rate in some areas reaches more than 50%. When using MODIS data, the unavailable range is larger, which extends south to Hubei, Hunan, and Guangxi. According to the monthly distribution (Fig. 7c), the parameters fitted by NLSM are not applicable in spring and winter, possibly because the PW in these two seasons is relatively low. Further, both the experimental and the extrapolation groups show that the unavailability rate when using MODIS data is larger than when using MISR data.

The application of the PSO algorithm in the M-Elterman mode has the advantage of global convergence; however, the initial value of each iteration is random, which leads to different experimental results. This indicates that the optimal parameter solution of different experimental simulations for the same station is not the only solution. In theory, if each station repeats the

iterative experiment of the PSO algorithm, the optimal solution can be selected from different experimental results, increasing the number and credibility of the available station.

Taking the simulation results using MODIS data as an example, we plot the spatial distribution of evaluation parameters KGE and R for the four PSO repeated experiments (Fig. 8). The results of the four experiments show that the better and more stably performing stations are mainly distributed in the North China Plain, Southwest China (including Tibet, western Sichuan, Yunnan, Guangxi, and Guangdong), and southern Xinjiang. However, there are still differences between several stations, and the inversion results of these stations in different batches of experiments are not in the same evaluation range. With MISR data, the repeated PSO experiment results show similar distribution characteristics (abridged).

According to the relatively stable, but slightly different distribution characteristics of the inversion results, we take KGE and R as the thresholds and select available stations from different experimental batches based on the results of the first experiment (i.e., the stations that meet the criteria are further selected from the stations not selected in the last experiment). The selection of stations in each experiment is shown in Table 2. With the increase in the number of experiments, the available stations in each evaluation range increased but the number of increased stations is less and less. Taking KGE & R > 0.3 as an example, in the first experiment, 413 stations could be selected when using MODIS
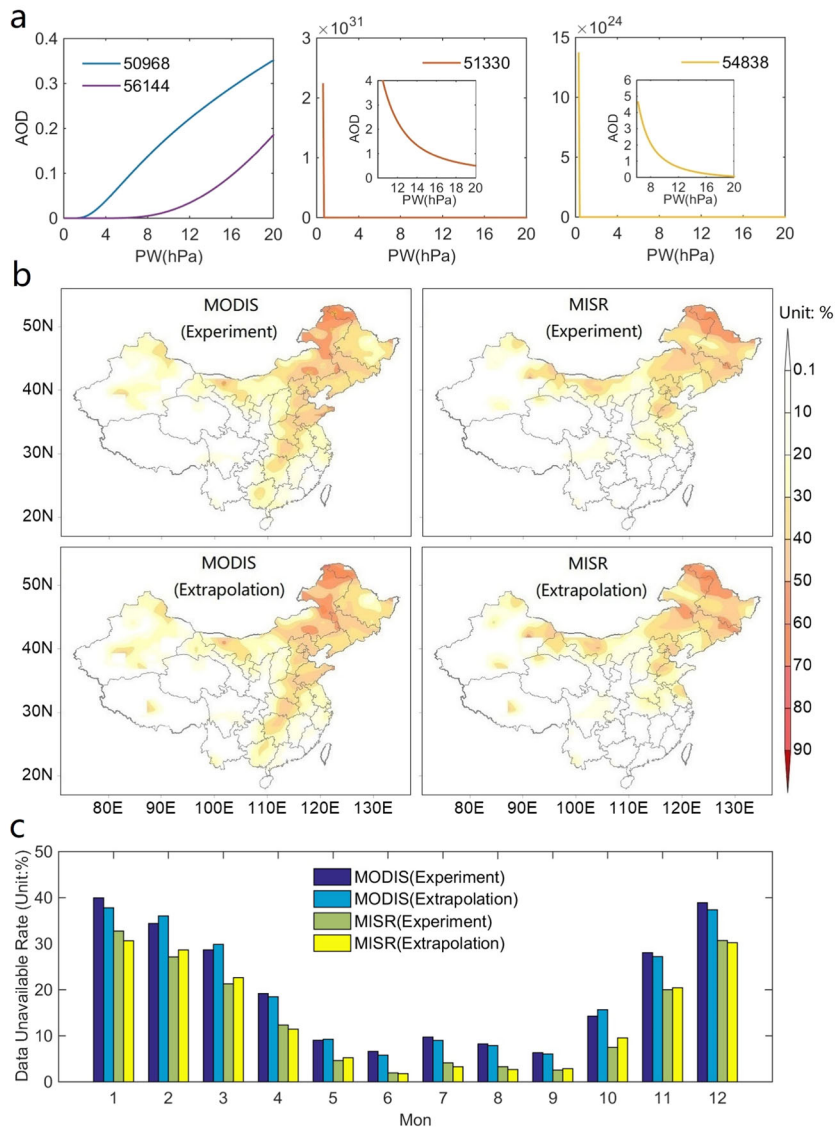
Fig. 7 Limitations of NLSM. a Variation of AOD with PW in different stations when using MISR data. **b** Spatial distribution and (**c**) Monthly distribution of the unavailable rate of inversion AOD.

data, while in the fourth experiment, 467 sites could be selected, which amounts to an increase of 54 in total. A total of 308 stations are selected after four experiments using MISR data, an increase of 56 in total. Table 2 shows that for a specific station, although the parameters fitted by different PSO algorithms are different each time, the difference between the simulated results and the real value is still relatively stable. Nevertheless, repeated experiments can still help increase the number of credible stations with more reliable simulation results. The final stations selected for the PSO algorithm in Fig. 2 are based on the scheme above discussed in this section.

This study improves the processing method of inconsistent recording in 661 meteorological stations in China, which improves the uniformity and reliability of visibility data collected prior to 1980 and subsequently employs them for AOD retrieval. The PSO algorithm is introduced into the AOD retrieval and applied in combination with the NLSM as the PSO-M-Elterman model. MODIS and MISR satellite data are employed as true values to retrieve AOD using visibility data. Both the scheme of visibility grade conversion and the PSO-M-Elterman model refine the evaluation results to the scale of a single station, and achieve good experimental results. Furthermore, the long-term change of AOD

in China is found to mainly occur in areas with a large population, which proves that the increase in AOD is caused by human activity rather than natural causes. The PSO-M-Elterman model can be generally applied to global AOD retrieval using monthly or daily visibility data to reconstruct long-term AOD data.

METHODS
Grade-distance conversion method of visibility

Visibility was recorded on a scale of ten grads before 1980: below 0.05; 0.05–0.2; 0.2–0.5; 0.5–1; 1–2; 2–4; 4–10; 10–20; 20–50 and above 50 (unit: km), and the actual visibility distance (unit: km) was directly recorded after 1980. Theoretically, each visibility level can correspond to any value within the corresponding distance interval, such as the median value of the interval. Since the interval of visibility distance corresponding to each visibility grade is not equally spaced, the use of median value instead of each visibility grade may result in unhomogeneity of data. Qin et al. developed a grade-distance conversion method:

Firstly, the actual observed visibility distance of all stations after 1980 is converted into the corresponding grade according to the 10 intervals above mentioned. (The time range is from 1980 to 2005 in Qin et al.'s research).
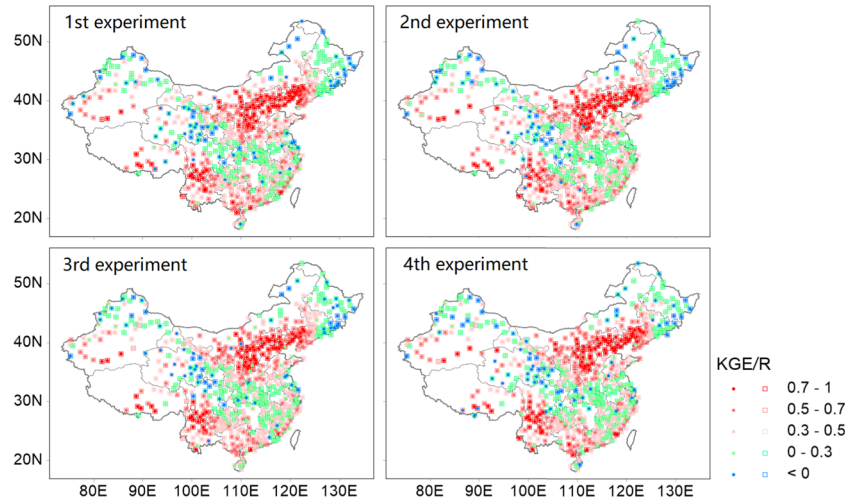
Fig. 8 Repeat experiment of PSO algorithm. Distribution of KGE and R in four repeated experiments of PSO algorithm.

Table 2. The comparison of the station numbers that can be selected for each time in PSO repeated experiment at different KGE and R value intervals.

Number of stations		KGE & $R > 0.7$	KGE & $R > 0.5$	KGE & $R > 0.3$	KGE & $R > 0$
MODIS	1st experiment	77	221	413	587
	2nd experiment	18	25	35	30
	3rd experiment	5	7	14	8
	4th experiment	1	2	5	8
	Total	101	255	467	633
MISR	1st experiment	4	55	252	520
	2nd experiment	0	20	35	32
	3rd experiment	0	8	17	18
	4th experiment	0	0	4	4
	Total	4	83	308	574

Then, calculate the average value of distance at each visibility grades according to the corresponding relationship in the first step.

Finally, the average distance of each grade is substituted as an estimated value for that grade before 1980.

It is worth noting that all stations share the same conversion scheme in Qin's method. However, China has a diverse geographical environment and observation conditions are complex at different stations. In addition, visibility also shows seasonal differences. The same transformation scheme obviously does not represent the individual differences of each station. Therefore, we improved it on the basis of Qin's method by respectively calculating the average value of visibility distance for 12 months at each station after 1980 (The actual time range involved in calculating the average distance of each grade are only from 1980 to1985 to eliminate the long-time trending signal of visibility). That is, each station has its own independent grade-distance conversion scheme for each month, and there are 661(stations) × 12(months) conversion schemes that are used for visibility grade data before 1980 in total.

M-Elterman model

M-Elterman model is developed on the basis of the original conversion method of visibility and extinction coefficient

According to the definition of Koschmieder's law[41]: the relation between visibility and the extinction coefficient $\sigma_{0.55}$ of 0.55 μm wavelength is:

$$V = \frac{3.912}{\sigma_{0.55}} \quad (5)$$

Assuming aerosols obey a Junge distribution:

$$n(r) = cr^{-(v^*+1)} \quad (6)$$

where c is a constant and r is the particle radius, and $v^* = 3$ does not change with height.

Under the standard surface air temperature (15 °C) and pressure (1013 hPa) conditions, the air molecular extinction coefficient is a constant 0.0116, then the aerosol extinction coefficient σ_λ and the atmospheric molecules σ_m at the height of z at the wavelength of λ can be expressed as:

$$\sigma_\lambda = \frac{N^A(z)}{N^A(0)}\left(\frac{3.912}{V} - 0.0116\right)\left(\frac{0.55}{\lambda}\right) \quad (7)$$

$$\sigma_m = 0.0116 - 0.00099z \quad (8)$$

Where $N^A(0)$ and $N^A(z)$ are the concentrations of aerosol particles at the surface and the altitude of z respectively, and the distribution of $N^A(z)$ with altitude proposed by McClatchey[42] can be expressed as:

$$N^A(z) = \begin{cases} 55\exp\left[-\frac{(z-5.5)}{H_1}\right] & (z \leqslant 5.5\,\text{km}) \\ 55 & (5.5\,\text{km} \leqslant z \leqslant 18\,\text{km}) \\ 55\exp\left[-\frac{(z-18)}{H_2}\right] & (z > 18\,\text{km}) \end{cases} \quad (9)$$

Then the relationship between the visibility V_z actually observed at the altitude of z and the visibility V revised to sea level can be represented as:

$$V_z = \frac{3.912}{0.0116 - 0.00099z + \left(\frac{3.912}{V} - 0.0116\right)e^{-\frac{z}{(0.886+0.222V)}}} \quad (10)$$

Thus, the AOD τ_λ can be obtained by integrating the extinction coefficient in the vertical direction:

$$\tau_\lambda = \left(\frac{3.912}{V} - 0.0116\right)\left(\frac{0.55}{\lambda}\right)^{2-v^*}\left[H_1\left(e^{-\frac{z}{H_1}} - e^{-\frac{5.5}{H_1}}\right) + 12.5e^{-\frac{z}{H_1}} + H_2e^{-\frac{z}{H_1}}\right] \quad (11)$$

where $H_1 = 0.886 + 0.0222V(km)$.

Formula (11) is the classical Elterman model of visibility inversion AOD. But it is sensitive to the geographical distribution of China, especially in coastal areas. Qiu[24] developed a parameterized model suitable for Chinese characteristics by comparing it with the optical depth detected through direct solar radiation data:

$$\tau_\lambda = \left(\frac{3.912}{V} - 0.0116\right)\left(\frac{0.55}{\lambda}\right)^{2-v^*}\left[H_1\left(e^{-\frac{z}{H_1}} - e^{-\frac{5.5}{H_1}}\right) + 12.5e^{-\frac{z}{H_1}} + H_2e^{-\frac{z}{H_1}}\right] \cdot f \quad (12)$$

$$f = \begin{cases} e^{(-0.32+0.02V_z)} & \text{for Northeast China} \\ e^{(0.42+0.0046P_w+0.015V_z)\cdot\exp\left(-0.0047\frac{V_z^2}{P_w}\right)} & \text{for other region} \end{cases} \quad (13)$$

In this model, the value of AOD depends on visibility V_z, water vapor pressure P_w, altitude z, wavelength, and Junge spectral parameter v^*. The revised coefficient f depends on the distribution of aerosol particle number density with height. $f < 1$ means that the attenuation of aerosol

particle number density with height is faster than that of Elterman model, corresponding to Northeast China; and $f > 1$ means that the attenuation of aerosol particle number density with height is slower than that of Elterman model, corresponding to the other region of China.

The M-Elterman model is based on Qiu's model[24] proposed by Wu et al.[20]. It adjusts the empirical constant as a variable parameter for each station:

$$AOD_i = F(V; a, b, c, d, e, f, g) \tag{14}$$

F in the formula is:

$$F = \left(\frac{3.912}{V} - 0.0116\right)\left(\frac{0.55}{\gamma}\right)^{2-V^*}\left[(e \cdot V + f)\left(\exp\left(-\frac{z}{e \cdot V + f}\right) - \exp\left(-\frac{5.5}{e \cdot V + f}\right)\right)\right.$$
$$\left. + 12.5\exp\left(-\frac{5.5}{e \cdot V + f}\right) + g \cdot \exp\left(-\frac{5.5}{e \cdot V + f}\right)\right] \cdot \exp\left((a + b \cdot p_w + c \cdot V_z)\exp\left(\frac{d \cdot V^2}{p_w}\right)\right) \tag{15}$$

AOD_i represents the inversion AOD at the ith station; a, b, c, d, e, f, and g are the parameters corresponding to each station.

The NLSM is employed to calculate each parameter and define DIF:

$$DIF = \sum_{Year_{start}}^{Year_{end}} (AOD - AOD_i)^2 = \sum_{Year_{start}}^{Year_{end}} (AOD - F(V; a, b, c, d, e, f, g))^2 \tag{16}$$

An independent set of parameters a, b, c, d, e, f, and g for each station can be obtained through iterative nonlinear equations. Then, the corresponding parameters can be applied to the year with the missing satellite data, and the AOD can be retrieved using visibility.

Particle swarm optimization algorithm

When using the PSO algorithm to optimize parameters, the objective function needs to be defined. Nash-Sutcliffe Efficiency (NSE) is usually used as the objective function:

$$NSE = 1 - \frac{\sum_k (S_k - O_k)^2}{\sum_k (O_k - O)^2} \tag{17}$$

where S_k and O_k represent the simulated value and observed value respectively, and O represents the observed average value. NSE is used to measure the degree of fit between observed and simulated values. The range of variation is from $-\infty$ to 1, and the closer to 1, the better the simulation performance.

The PSO algorithm also depends on the parameters of the algorithm itself: the number of particle swarms, the position range and the velocity range of each particle. For the parameters to be optimized, its range can be changed to $[-1, 1]$ through the following mapping:

$$x = \frac{2y - (R_{max} + R_{min})}{(R_{max} - R_{min})} \tag{18}$$

where, y is the ture value of the parameter, and R_{max} and R_{min} are the actual maximum and minimum values of the parameter.

The mathematical representation of the PSO algorithm is as follows:

Suppose that to optimize an n-dimensional problem, m particles are selected, where the position and velocity vector of the ith particle can be expressed as:

$$x_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \tag{19}$$

$$v_i = (v_{i1}, v_{i2}, \ldots, v_{in}) \tag{20}$$

The updated position and velocity of the ith particle are respectively expressed as:

$$v_{in}^{N+1} = wv_{in}^N + c_1 r_1 (p_{in}^N - x_{in}^N) + c_2 r_2 (G_n^N - x_{in}^N) \tag{21}$$

$$x_{in}^{N+1} = x_{in}^N - v_{in}^N \tag{22}$$

where N represents the number of iterations, w represents the weight of inertia, c_1 and c_2 represents the acceleration constant, and is the weight coefficient for particles to track their own historical optimal value, which is used to measure the recognition of particles themselves. r_1 and r_2 are random numbers between $[0, 1]$, and p_i and G_n respectively represent the local and global optimal position searched by the ith particle. They can

be respectively expressed as:

$$p_i = (p_{i1}, p_{i2}, \ldots, p_{in}) \tag{23}$$

$$G_n = (p_{g1}, p_{g2}, \ldots, p_{gn}) \tag{24}$$

$$g = \min_{1 \leqslant i \leqslant n} [F(p_i)] \tag{25}$$

where g is the lowest value of the target function, and F is the target function.

DATA AVAILABILITY

Meteorology data over 661 Chinese meteorological sites used for the present study was from Daily data set of Surface climatic data in China V3.0 and provided by the China Meteorological Administration(http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_MUL_DAY_V3.0/). AOD data for 2001–2014 were from NASA (National Aeronautics and Space Administration) Goddard Earth Sciences Data and Information Services Center: Multiangle Imaging Spectro Radiometer (MISR) daily Level 3 Component Global Aerosol Product (https://asdc.larc.nasa.gov/data/MISR/MIL3DAE.004/) and Moderate Resolution Imaging Spectroradiometer (MODIS) Level-3 aerosol products of the Terra satellite (https://nrt3.modaps.eosdis.nasa.gov/allData/61/MODAODHD/). The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

CODE AVAILABILITY

The computer code of Particle Swarm Optimization algorithm in this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Boucher, O. et al. *Clouds and Aerosols, in Climate Change 2013*. Vol. 7 (Cambridge University Press, 2013).
2. Chin, M. et al. Multi-decadal aerosol variations from 1980 to 2009: a perspective from observations and a global model. *Atoms. Chem. Phys.* **14**, 3657–3690 (2014).
3. Xin, J. et al. Aerosol optical depth (AOD) and Ångström exponent of aerosols observed by the Chinese sun hazemeter network from August 2004 to September 2005. *J. Geophy. Res-Atmos.* **112**, 1703–1711 (2007).
4. Wang, L. et al. Evaluation of the MODIS aerosol optical depth retrieval over different ecosystems in China during EAST-AIRE. *Atmos. Environ.* **41**, 7138–7149 (2007).
5. Li, Z. et al. Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective. *Ann. Geophys-Germany.* **27**, 2755–2770 (2009).
6. Li, Z. et al. East asian studies of tropospheric aerosols and their impact on regional climate (EAST-AIRC): an overview. *J. Geophys. Res-Atmos.* **116**, D00K34 (2011).
7. Xia, X. et al. Ground-based remote sensing of aerosol climatology in China: aerosol optical properties, direct radiative effect and its parameterization. *Atmos. Environ.* **124**, 243–251 (2016).
8. Zhang, S., Wu, J., Fan, W., Yang, Q. & Zhao, D. Review of aerosol optical depth retrieval using visibility data. *Earth-Sci. Rev.* **200**, 102986 (2020).
9. Lau, K. M., Kim, M. K. & Kim, K. M. Asian summer monsoon anomalies induced by aerosol direct forcing: the role of the Tibetan Plateau. *Clim. Dynam.* **26**, 855–864 (2006).
10. Bollasina, M. A., Yi, M. & Ramaswamy, V. Anthropogenic aerosols and the weakening of the South Asian summer monsoon. *Science.* **334**, 502–505 (2011).
11. Song, F., Zhou, T. & Yun, Q. Responses of East Asian summer monsoon to natural and anthropogenic forcings in the 17 latest CMIP5 models. *Geophy. Res. Lett.* **41**, 596–603 (2014).
12. Zhang, L., Liao, H. & Li, J. Impacts of Asian summer monsoon on seasonal and interannual variations of aerosols over eastern China. *J. Geophys. Res-Atmos.* 115 (2010).
13. Shen, Y. B., Shen, Z. B. & Wang, W. F. Atmospheric aerosol optical thickness and dusty weather in northern China in spring of 2001. *Plateau. Meteor. (in Chinese).* **22**, 185–190 (2003).
14. Hu, B., Wang, Y. S. & He, X. X. Variation properties of earth's surface solar radiation during a strong dust storm in Beijing 2004. *Clim. Environ. Res. (in Chinese).* **71**, 292–306 (2005).

15. Ke, Z. D. & Tang, J. An observation study of the scattering properties of aerosols over Shangdianzi, Beijing. *Atmos. Sci. (in Chinese)*. **31**, 553–559 (2007).

16. Qiu, J. A method to determine atmospheric aerosol optical depth using total direct solar radiation. *J. Atmos. Sci.* **55**, 744–757 (1998).

17. Wang, K., Dickinson, R. E. & Liang, S. Clear sky visibility has decreased over land globally from 1973 to 2007. *Science*. **323**, 1468–1470 (2009).

18. Chen, L. X. et al. Variation of atmospheric aerosol optical depth and its relationship with climate change in China east of 100°E over the last 50 years. *Theor. Appl. Climatol.* **96**, 191–199 (2008).

19. Qin, S. G., Shi, G. Y. & Chen, L. Long-term variation of aerosol optical depth in China based on meteorological horizontal visibility observations. *Atmos. Sci. (in Chinese)*. **34**, 449–456 (2010).

20. Wu, J. et al. Improvement of aerosol optical depth retrieval using visibility data in China during the past 50 years. *J. Geophys. Res-Atmos.* **119**, 13,370–313,387 (2014).

21. Zhang, Z. et al. Aerosol optical depth retrieval from visibility in China during 1973–2014. *Atmos. Environ.* **171**, 38–48 (2017).

22. Boers, R. et al. Observations and projections of visibility and aerosol optical thickness (1956–2100) in the Netherlands: impacts of time-varying aerosol composition and hygroscopicity. *Environ. Res. Lett.* **10**, 015003 (2015).

23. Retalis, A. et al. Comparison of aerosol optical thickness with in situ visibility data over Cyprus. *Nat. Hazards. Earth. Sys.* **10**, 421–428 (2010).

24. Qiu, J. H. & Lin, Y. R. A parameterization model of aerosol optical depths in China. *J. Meteorol. Res-PRC.* **59**, 368–372 (2001).

25. Lin, J., Donkelaar, A. V., Xin, J., Che, H. & Wang, Y. Clear-sky aerosol optical depth over East China estimated from visibility measurements and chemical transport modeling. *Atmos. Environ.* **95**, 258–267 (2014).

26. Lin, J. & Li, J. Spatio-temporal variability of aerosols over East China inferred by merged visibility-GEOS-Chem aerosol optical depth. *Atmos. Environ.* **132**, 111–122 (2016).

27. Kennedy, J. & R, E. Particle Swarm Optimization. In *Proc IEEE Int Conf on Neurai Networks*. 1942–1948 (1995).

28. Schmitt, M. & Wanka, R. Particle swarm optimization almost surely finds local optima. *Theor. Comput.Sci.* **561**, 57–72 (2015).

29. Eberhart, R. C. & Shi, Y. Particle swarm optimization: development, applications and resources. In *Evolutionary Computation*. 81–86 (2001).

30. Gill, M. K., Kaheil, Y. H., Khalil, A., Mckee, M. & Bastidas, L. Multiobjective particle swarm optimization for parameter estimation in hydrology. *Water. Resour. Res.* **42**, 257–271 (2006).

31. Chau, K. W. Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. *J. Hydrol.* 329, (2006).

32. Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **377**, 80–91 (2009).

33. Zhang, X. Y. et al. Atmospheric aerosol compositions in China: spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols. *Atmos. Chem. Phys.* **11**, 26571–26615 (2012).

34. Engelhart, G. J., Hildebrandt, L., Kostenidou, E., Mihalopoulos, N. & Donahue, N. M. Water content of aged aerosol. Atmos. Chem. Phys. 11, (2011).

35. Liu, P. F., Zhao, C. S., Göbel, T., Hallbauer, E. & Wiedensohler, A. Hygroscopic properties of aerosol particles at high relative humidity and their diurnal variations in the North China Plain. *Atmos. Chem. Phys.* **11**, 3479–3494 (2011).

36. Hu, D. W. et al. Hygroscopicity of inorganic aerosols: size and relative humidity effects on the growth factor. *Aerosol. Air. Qual. Res.* **10**, 255–264 (2010).

37. Howell, S. G., Clarke, A. D., Shinozuka, Y. & Kapustin, V. Influence of relative humidity upon pollution and dust during ACE-Asia: Size distributions and implications for optical properties. *J.Geophys. Res-Atmos.* **111**, D06205 (2006).

38. Cheng, Y. F. et al. Relative humidity dependence of aerosol optical properties and direct radiative forcing in the surface boundary layer at Xinken in Pearl River Delta of China: An observation based numerical study. *Atmos. Environ.* **42**, 6373–6397 (2008).

39. Twohy, C. H., Coakley, J. A. & Tahnk, W. R. Effect of changes in relative humidity on aerosol scattering near clouds. *J.Geophys. Res-Atmos* 114, (2009).

40. Bian, H., Chin, M., Rodriguez, J. M., Yu, H. & Strahan, S. Sensitivity of aerosol optical thickness and aerosol direct radiative effect to relative humidity. *Atmos. Chem. Phys. Discuss.* **8**, 23785–22386 (2008).

41. Koschmieder, H. Therie der horizontalen sichtweite.*Beitr Phys.d.freien Atm* **12**, 33–55 (1924).

42. Mcclatchey, R. A. Optical properties of the atmosphere. *Afcrl. Environ. Res.* **108**, 3048 (1971).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Jian Wu designed the study and led the collaboration. Shuang Zhang performed the analysis, wrote the paper. Qidong Yang contributed model code. Deming Zhao, Wenxuan Fan, and Jinchuan Zhao participated in the revision of the article. Cheng Shen provided technical support for the figures. All authors contributed to the discussions in the paper.

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/s41612-021-00207-5.

Correspondence and requests for materials should be addressed to Jian Wu.

Reprints and permission 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.