

# SCIENTIFIC REPORTS

OPEN

## Combining Positive Matrix Factorization and Radiocarbon Measurements for Source Apportionment of PM<sub>2.5</sub> from a National Background Site in North China

Xiaoping Wang<sup>1,2</sup>, Zheng Zong<sup>2</sup>, Chongguo Tian<sup>2</sup>, Yingjun Chen<sup>3</sup>, Chunling Luo<sup>1</sup>, Jun Li<sup>1</sup>, Gan Zhang<sup>1</sup> & Yongming Luo<sup>2</sup>

To explore the utility of combining positive matrix factorization (PMF) with radiocarbon (<sup>14</sup>C) measurements for source apportionment, we applied PM<sub>2.5</sub> data collected for 14 months at a national background station in North China to PMF models. The solutions were compared to <sup>14</sup>C results of four seasonally averaged samples and three outlier samples. Comparing the most readily interpretable PMF solutions and <sup>14</sup>C results revealed that PMF modeling was well able to capture the source patterns of PM<sub>2.5</sub> with two and three irrelevant source classifications for the seasonal and outlier samples. The contribution of sources that could not be classified as either fossil or non-fossil sources in the PMF solution, and the errors between the modeled and measured concentrations weakened the effectiveness of the comparison. Based on these two factors, we developed an index for selecting the most suitable <sup>14</sup>C measurement samples for combining with the PMF model. Then we examined the potential for coupling PMF modeling and <sup>14</sup>C data with a constrained PMF run using the <sup>14</sup>C data as a priori information. The restricted run could provide a more reliable solution; however, the PMF model must provide a flexible dialog to input the priori restrictions for executing the constraint simulation.

Atmospheric fine particles with aerodynamic diameters <2.5 μm (PM<sub>2.5</sub>) are recognized as key pollutants associated with increased rates of mortality and morbidity for respiratory and cardiovascular diseases<sup>1</sup>. Reliable source apportionment of PM<sub>2.5</sub> is vital to design effective programs and strategies to reduce PM<sub>2.5</sub> concentrations in ambient air. However, source apportionment is complicated, because source identification and quantification generally cannot be directly measured or monitored<sup>2</sup>. These limitations have encouraged the development of novel approaches for the source apportionment of PM<sub>2.5</sub><sup>3</sup>.

Receptor modeling is one method used for the source apportionment of PM<sub>2.5</sub><sup>4</sup>. Widely used receptor models include principal component analysis/multiple linear regression, unmix, positive matrix factorization (PMF), and chemical mass balance<sup>3–5</sup>. These models apportion PM<sub>2.5</sub> to sources by decomposing a matrix of speciated PM<sub>2.5</sub> sample data collected from one or more receptor sites. In general, these models provide interpretable results, but the reliability of the results is not guaranteed<sup>3</sup>. To enhance the confidence level of the source apportionments or evaluate model uncertainties, receptor models can be combined with other methods. There are two main types of combinations. The first type uses a dataset of PM<sub>2.5</sub> speciation measurements to drive different receptor

<sup>1</sup>State Key Laboratory of Organic Geochemistry, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou, 510640, China. <sup>2</sup>Key Laboratory of Coastal Zone Environmental Processes and Ecological Remediation, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, 264003, China. <sup>3</sup>Key Laboratory of Cities' Mitigation and Adaptation to Climate Change in Shanghai (CMA), College of Environmental Science and Engineering, Tongji University, Shanghai, 200092, China. Correspondence and requests for materials should be addressed to C.T. (email: [cgtian@yic.ac.cn](mailto:cgtian@yic.ac.cn))

models (or other types of models) for the source apportionment. For example, Song *et al.* used two receptor models and Bove *et al.* used one receptor model and a chemical transport model to apportion PM<sub>2.5</sub> to sources<sup>6,7</sup>. Balachandran *et al.* developed specific source profiles using Bayesian-based ensemble averaging of source impacts from three receptor models and one chemical transport model to apportion PM<sub>2.5</sub> to sources<sup>2</sup>. The second type of combination employs a suite of different PM<sub>2.5</sub> datasets to drive a receptor model. For example, Xie *et al.* used a PMF model with four different PM<sub>2.5</sub> speciation datasets to evaluate the utility and consistency of the source apportionments<sup>8</sup>. Tao *et al.* loaded a single or combination of three biomass tracers with other species within PM<sub>2.5</sub> into a PMF model to apportion PM<sub>2.5</sub> to sources and to assess the uncertainties of the source contributions<sup>9</sup>. It should be noted that these combinations could be used to compare the coherence and uncertainty among different results, but cannot guarantee their reliability.

Carbon, nitrogen, and sulfur isotope measurements have been used to identify the sources of the respective PM<sub>2.5</sub> components<sup>10–12</sup>, which could provide more reliable source signatures than the methods mentioned above. Radiocarbon (<sup>14</sup>C) measurements can be used to unambiguously quantify the contribution of fossil and non-fossil sources to carbonaceous components of PM<sub>2.5</sub> because <sup>14</sup>C becomes depleted in fossil sources due to their aging (half-life, 5,730 years), whereas non-fossil sources contain similar <sup>14</sup>C levels as atmospheric carbon dioxide (CO<sub>2</sub>)<sup>13–17</sup>. However, there have been few attempts to combine <sup>14</sup>C-derived source signatures with receptor models for the source apportionment of PM<sub>2.5</sub><sup>18,19</sup>.

We recently performed a preliminary exploration of combining PMF modeling and <sup>14</sup>C measurements for the source apportionment of wintertime PM<sub>2.5</sub><sup>18</sup>. To gain further insight into this method, we collected PM<sub>2.5</sub> samples for 14 months at a national background atmospheric monitoring station on Tuoji Island, North China. Performing source apportionment at a background site could offer the opportunity to identify regional sources of PM<sub>2.5</sub> in North China, where the adverse effects of PM<sub>2.5</sub> on public health have triggered both public alarm and official concern<sup>20</sup>. Moreover, these sample data are appropriate for assessing PMF performance and the utility of combining PMF and <sup>14</sup>C data to apportion PM<sub>2.5</sub> to sources, because there are no obvious source emissions near the sampling site, and airborne pollutants are mixed and transformed well before reaching the sampling site. The main objectives of this study were: (1) to assess the capability of using PMF modeling for the source apportionment of PM<sub>2.5</sub> by comparing PMF and <sup>14</sup>C results; (2) to develop an index to select suitable <sup>14</sup>C measurement samples for the most effective combination of <sup>14</sup>C measurements with PMF to apportion PM<sub>2.5</sub> to sources due to the high cost and complex pretreatment of <sup>14</sup>C measurements; and (3) to propose additional improvements to PMF modeling software for coupling PMF and <sup>14</sup>C data to increase the reliability of PM<sub>2.5</sub> source apportionments.

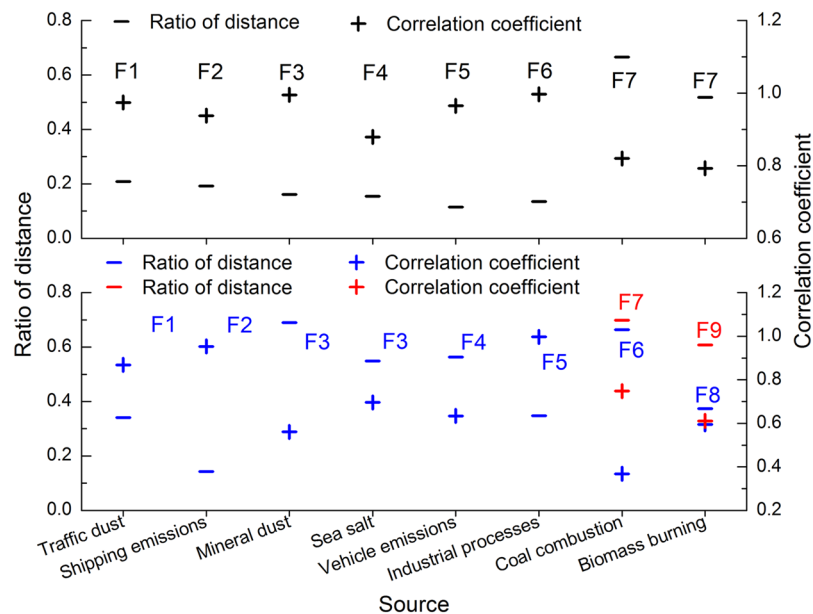
## Results and Discussion

**General characteristics of PM<sub>2.5</sub>.** During the entire sampling period, the mean ± standard deviation of the PM<sub>2.5</sub> mass concentration was 57.7 ± 36.9 μg m<sup>-3</sup> (range, 8.93–144 μg m<sup>-3</sup>). The dominant PM<sub>2.5</sub> species were SO<sub>4</sub><sup>2-</sup>, NO<sub>3</sub><sup>-</sup>, OC, NH<sub>4</sub><sup>+</sup>, EC, K<sup>+</sup>, and Fe, accounting for 17.3 ± 7.96%, 10.3 ± 6.48%, 7.98 ± 3.08%, 4.81 ± 2.27%, 4.12 ± 3.61%, 1.13 ± 0.617%, and 0.89 ± 0.64%, respectively, of the PM<sub>2.5</sub> mass concentration. The seasonal averages of PM<sub>2.5</sub> and most species concentrations were markedly higher in the spring of 2012 than the other four seasons (Supporting Information [SI] Text S1, Table S1, and Figures S1–S3).

**PMF modeling.** In the BMRs with six to nine factors, all of the BS-DISP estimations showed that the largest decreases in Q(robust) were <1% of the Q(robust) of the corresponding BMR (SI Figure S4), suggesting that the results of the four BMRs could be considered the global optimum solutions for the corresponding number of factors. The F<sub>peak</sub> model runs with strengths of -1.0, 1.0, and 1.5 showed increases in Q > 5% of Q(robust) from the four respective BMRs (SI Figure S5), indicating that these F<sub>peak</sub> solutions should be removed from further consideration. However, the F<sub>peak</sub> runs with strengths of -0.5 and 0.5 showed increases in Q < 5% of Q(robust) of the corresponding BMRs (SI Figure S5), suggesting that these F<sub>peak</sub> solutions were acceptable for further analysis. Positive/negative F<sub>peak</sub> strengths sharpened/smear and smear/sharpened the two solution matrices, respectively<sup>21</sup>. The comparable increases in Q values and opposite transformations of the two matrices suggested that there were no significant rotational effects of the two strengths of the BMR solutions. Among the four model experiments, the BMR with eight factors (BMR-8) showed the most significant declines in Q(true), Q(robust), and the ratio of Q(true)/Q(exp) (SI Figures S6 and S7). These declines suggested that the source apportionments of BMR-8 were more appropriate than those of BMR-6, BMR-7, and BMR-9.

The eight source factors in BMR-8 were identified based on the dominant species and prominent contributions of a source to each species on a specified sampling day. SI Text S2 and Figures S8–S14 elaborate on the recognition processes used in this study. The eight identified sources were traffic dust, shipping emissions, mineral dust, sea salt, vehicle emissions, industrial processes, coal combustion, and biomass burning. The total and seasonal contributions (%) of the eight sources to PM<sub>2.5</sub> were calculated, and results are shown in SI Figure S15. Among these sources, biomass burning, shipping emissions, and coal combustion had the largest contributions, accounting for 27.5%, 17.5%, and 16.5%, respectively, of the PM<sub>2.5</sub> mass concentration during the entire sampling period. These sources were followed by mineral dust (14.8%), vehicle emissions (10.4%), sea salt (4.8%), traffic dust (4.6%), and industrial processes (3.9%). The back trajectory, fire counts, and navigation activity information supported these identified sources. Details are presented in SI Text S3 and Figures S11–S19.

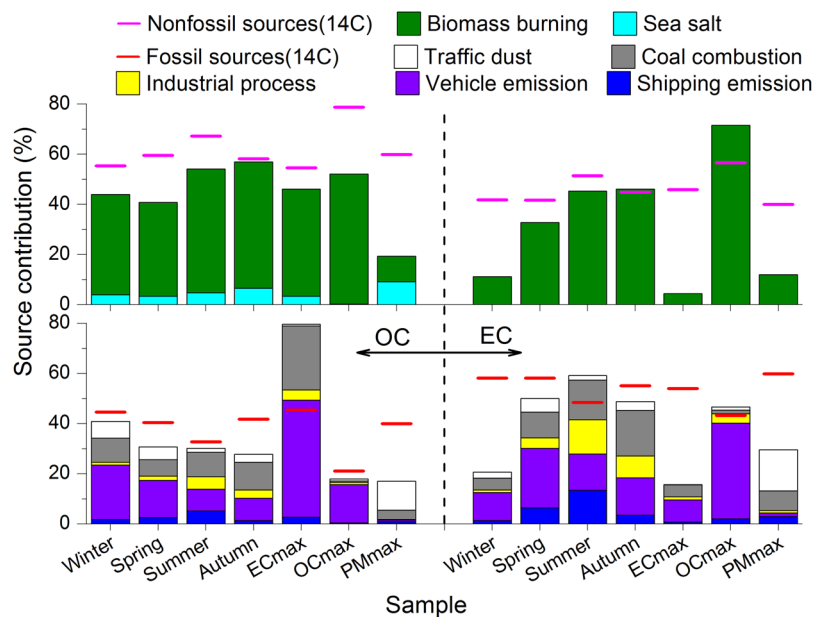
To quickly identify the sources modeled by BMR-7 and BMR-9, the distances between their source profiles and that of BMR-8, and the Pearson correlations of their contribution time series with those of BMR-8 were calculated. Overall, most of the sources in BMR-8 matched the factors in BMR-7 and BMR-9 well. Six sources in BMR-7 were closely related to six sources identified by BMR-8, while one factor merged the two remaining sources (coal combustion and biomass burning) in BMR-8, characterized by their shortest distances and highest correlation coefficients (Fig. 1, detailed data are listed in SI Tables S2 and S3). This combined source had a lower contribution fraction to PM<sub>2.5</sub> than the sum of the two separate sources (coal combustion and biomass burning)



**Figure 1.** Ratios of the shortest distance between the standardized seven-factor (top panel) and nine-factor (bottom panel) source profiles from BMR-7 and BMR-9 and those from BMR-8 to the averages of their respective distance, and the Pearson correlation coefficients of the contribution time series related to the source profile with the shortest distance (labeled with the same color as that for the distance). For BMR-9, the source profile shows the shortest distance with two source profiles from BMR-8 (F3). Coal combustion and biomass burning identified in BMR-8 were each related to two different source profiles BMR-9.

in BMR-8 (SI Figure S18). Among the eight sources identified by BMR-8, traffic dust, shipping emissions, vehicle emissions, and industrial processes matched the factors in BMR-9 well, characterized by their shortest distances and high correlation coefficients (Fig. 1, detailed data are listed in SI Tables S4 and S5). Both mineral dust and sea salt had the shortest distances and highest correlation coefficients with one factor in BMR-9, indicating that these two sources were combined in BMR-9 (Fig. 1). Also, coal combustion and biomass burning identified by BMR-8 each matched two separate factors in BMR-9. The two factors in BMR-9 related to coal combustion were industrial coal and domestic coal because their contribution time series correlated more with those of industrial processes and biomass burning, respectively (SI Table S5). Similarly, the two factors in BMR-9 associated with biomass burning were domestic biofuel and open biomass burning (SI Table S5). This source separation resulted in significant increases in the contribution of biomass burning and marked decreases in the contribution of coal combustion (SI Figure S19). We did not use the differences in the source contributions modeled by the three BMRs to determine the best solution among the three models. In BMR-7, coal combustion and biomass burning were inappropriately combined as a hybrid source, because they represent different types of emission sources and different primary contributors of  $PM_{2.5}$ , suggesting that BMR-7 did not provide the optimal solution. This was supported by the initial mathematical analysis, which showed that BMR-8 provided the most physically interpretable results.

**PMF performance assessment.** The model capacity was assessed by comparing the source contributions modeled by BMR-8 and BMR-9 with the  $^{14}C$  results. We did not consider BMR-7 for this evaluation, because the model included a source that combined fossil carbon (coal combustion) and non-fossil carbon (biomass burning) sources. According to the source types identified by BMR-8, coal combustion, industrial processes, vehicle emissions, and shipping emissions were ranked as fossil carbon sources, while sea salt as a marine biogenic source was merged with biomass burning as a non-fossil source. Mineral dust and traffic dust were not considered in this classification, because they could not be apportioned quantitatively into fossil and non-fossil sources. Figure 2 shows the contributions of the non-fossil and fossil sources categorized from the BMR-8 results and the corresponding  $^{14}C$  results for the seven samples. The source contribution from the BMR-8 results was calculated based on Equation (7) in the section of methods. For the four seasonal samples, from the winter of 2011 to autumn of 2012, the results of the non-fossil sources classified from BMR-8 accounted for 44.0%, 40.9%, 54.2%, and 57.0% of OC concentrations and 11.3%, 32.8%, 45.3%, and 46.2% of EC concentrations, respectively. Correspondingly, the categorized BMR-8 results showed that 34.3%, 25.6%, 28.6%, and 24.7% of OC and 18.4%, 44.7%, 57.4%, and 45.4% of EC were attributed to fossil sources, respectively. Most of the source contributions were lower than the corresponding  $^{14}C$  fractions. The underestimations were not considered to indicate inadequate capacity of the PMF model, because mineral dust and traffic dust were not taken into account for the comparison. In contrast, the overestimation of the source contributions classified from the PMF results could be considered irrelevant source apportionments of the PMF model. There were two source contribution overestimations of 8.9% and 1.3% for a fossil source of EC in summer and a non-fossil source of EC in autumn, respectively. This indicated



**Figure 2.** Comparison of the source apportionment of OC and EC classified from the PMF results and determined from the  $^{14}\text{C}$  measurements. Biomass burning and sea salt identified by PMF were combined as a non-fossil source, while coal combustion, industrial processes, vehicle emissions, and shipping emissions were merged as fossil sources for the comparison. Mineral dust was not included because it was considered a hybrid of non-fossil and fossil sources, and traffic dust was included as an additional fossil source in the comparison.

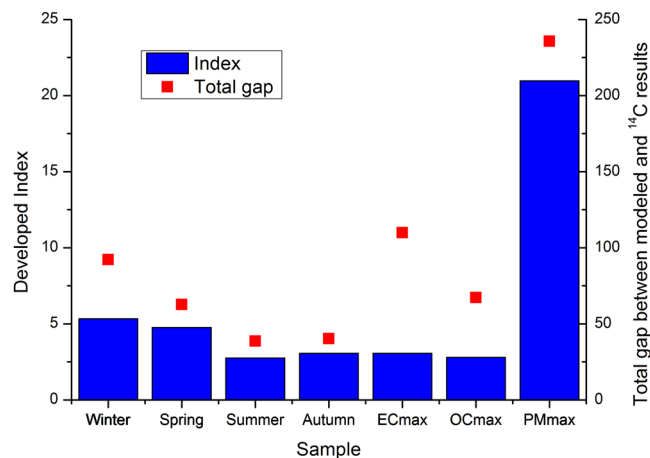
that non-fossil sources were irrelevantly classified as fossil sources in the former case and fossil sources were categorized inappropriately as non-fossil sources in the latter. Traffic dust is thought to contain a higher fraction of fossil carbon than mineral dust<sup>22</sup>. Therefore, to further estimate PMF performance, we classified traffic dust as an additional fossil carbon source and re-compared the source contributions. The contribution of vehicle dust was added to the stack columns in Fig. 2. After this addition, the gaps in the source contributions of BMR-8 and  $^{14}\text{C}$  results decreased significantly, while the irrelevant source classifications increased non-significantly, indicating that the PMF model captured the primary source patterns of  $\text{PM}_{2.5}$  well. SI Figure S20 displays the comparison of the  $^{14}\text{C}$  and PMF results calculated based on Equation (6) in the section of methods. The comparison shows a similar pattern as that in Fig. 2. This similarity indicated that the PMF model reproduced the seasonally measured concentrations well, and the differences in the source contributions modeled by BMR-8 and the  $^{14}\text{C}$  results were attributed to the effects of mineral dust rather than errors between the modeled and measured concentrations.

A similar comparison to that described above was performed for OC and EC in the outlier samples. Regardless of whether traffic dust was considered a fossil source, only three overestimations were observed, among which the two most marked overestimations were the contribution of fossil sources to OC (33.7% and 34.3% without and with consideration of vehicle dust as a fossil source) in the sample with the highest EC concentration and the contribution of non-fossil sources to EC (14.8%) in the sample with the highest OC concentration. In addition, a slight overestimation was observed for the contribution of fossil fuel combustion to EC (2.2% and 3.4% without and with consideration of vehicle dust as a fossil carbon source) in the sample with the highest OC concentration (Fig. 2). The comparison based on Equations (6) and (7) showed that the differences in source contributions for the samples with the highest OC and EC concentrations were the result of the contributions of both mineral dust and model error, while the difference in the sample with the highest  $\text{PM}_{2.5}$  concentration was mainly driven by mineral dust.

We compared the source contributions modeled by BMR-9 with the  $^{14}\text{C}$  results. For this comparison, domestic coal, industrial coal, industrial processes, vehicle emissions, and shipping emissions identified from BMR-9 were ranked as fossil sources, whereas domestic biomass burning and open biomass burning were classified as non-fossil sources. Traffic dust and the hybrid mineral dust–sea salt source were not considered. There were more overestimations in this comparison than that of BMR-8 (SI Figure S21), suggesting that the BMR-8 model provided a more reliable source apportionment than BMR-9.

**Index development.** We calculated the contribution fractions of non-fossil and fossil sources classified from the BMRs based on the corresponding measured concentrations, as shown in Equation (7) in the section of methods. The measured concentration can be replaced by the corresponding modeled concentration, and Equation (7) can be rewritten as:

$$R_{ij} = \sum_{k=1}^n g_{ik} f_{kj} / \left( \sum_{k=1}^p g_{ik} f_{kj} + e_{ij} \right), \quad (1)$$



**Figure 3.** The developed index (dimensionless) used to select for <sup>14</sup>C measurement and the total gap between the source contributions from the PMF and <sup>14</sup>C results (%).

where  $e$  is the error between the modeled and measured concentrations, and the other parameters are the same as those in Equation (6). According to the source classification of non-fossil and fossil sources, the sum of the two types of source contributions can be expressed as:

$$R_{ij} = \left( Cnf_{ij} + Cff_{ij} \right) / \left( Cnf_{ij} + Cff_{ij} + Cnc_{ij} + e_{ij} \right), \quad (2)$$

where  $Cnf$ ,  $Cff$ , and  $Cnc$  are the concentrations of the non-fossil, fossil, and unclassified source contributions, respectively.  $e$  indicates the modeled error similar to that in Equation (1). Unlike the <sup>14</sup>C results, the contribution fractions of non-fossil and fossil sources categorized from the PMF solution cannot add up to 1 because of the effects of unclassified sources and the error between the modeled and measured concentrations of OC and EC. The effects ( $eff$ ) can be considered the fraction of the sum of concentrations from unclassified source contributions and the error between the modeled and measured concentrations as:

$$eff_{ij} = \left( Cnc_{ij} + e_{ij} \right) / \left( Cnf_{ij} + Cff_{ij} + Cnc_{ij} + e_{ij} \right) = \left( Cnc_{ij} + e_{ij} \right) / X_{ij}, \quad (3)$$

where the symbols have the same definitions as those in Equations (3) and (5).

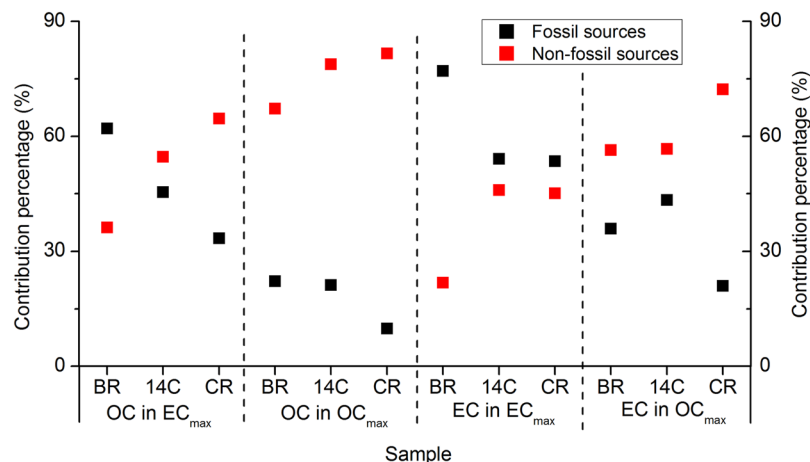
Large  $eff$  values generally resulted in greater underestimations of the contributions of sources categorized from the PMF results. These underestimations cannot be directly used to assess PMF model performance, but suggest a reduction in the comparison effectiveness. <sup>14</sup>C is usually measured for only a few independent samples due to the high cost and complicated pretreatment. This raises the question of how to select <sup>14</sup>C measurement samples to yield a more valid assessment of model performance and better combination of the PMF model with <sup>14</sup>C measurements. One feasible method is to develop an index to select the most efficient <sup>14</sup>C measurement samples for the PMF assessment. Based on the analysis as mentioned earlier, we developed the following index, which is dimensionless:

$$Index_j = \sum_{i=1}^n \left( \frac{Cnc_{ij} + e_{ij}}{X_{ij}} \right) \times \sum_{k=1}^m \left( \frac{Cnc_{kj} + e_{kj}}{X_{kj}} \right), \quad (4)$$

where the symbols have the same definitions as those in Equations (1), (2) and (7), and  $n$  and  $m$  represent the number of all species used for the PMF model and the number of carbonaceous species, respectively.

We calculated the index and total gaps in source contributions between the PMF and <sup>14</sup>C results for each sample shown in Fig. 2, and results are presented in Fig. 3. Among the seven samples, that with the highest PM<sub>2.5</sub> concentration had the largest total gap, which was attributed to the dominant contributions of mineral dust to PM<sub>2.5</sub> (69% for OC and 66% for EC). Its irrelevance was characterized well by its high index value. The average summer sample had the smallest total gap, which was in agreement with its lowest index value, indicating that it was the best option among the seven samples. In addition, the index values of each sample were calculated using the data from BMR-7, BMR-8, and BMR-9. The variation in the indices presented similar trends as shown in SI Figure S22, indicating the stability of the developed index. This suggested that <sup>14</sup>C measurement samples could be selected after several pre-simulations, rather than the final simulation, allowing for excellent performance of the assessment of the model's capacity and the combination of the PMF model with <sup>14</sup>C measurements. The stability of the index was attributed to similar trends in the errors between the modeled and measured PM<sub>2.5</sub>, OC, and EC concentrations and the modeled PM<sub>2.5</sub> concentrations contributed by mineral dust and traffic dust among BMR-7, BMR-8, and BMR-9 (SI Figures S23–S25).

**Potential coupling between PMF and <sup>14</sup>C measurement.** The relationship between the PMF and <sup>14</sup>C measurements discussed above was determined by a comparison, and suggested their independence. Further



**Figure 4.** Contributions of fossil and non-fossil sources to OC and EC in the base model run (BR), constrained model run (CR), and  $^{14}\text{C}$  measurements (14C).

exploration focused on whether the  $^{14}\text{C}$  results could be coupled to the PMF as *a priori* information to provide more reliable model results using a constrained PMF model. According to the previous knowledge, the  $^{14}\text{C}$  results can be used to build constraint equations that specify a range of the total contribution of several sources (factors) to either OC or EC (species) in an individual or incorporative samples. For instance, the total contribution of coal combustion, industrial processes, vehicle emissions, and shipping emissions to OC in the sample with the highest EC concentration was less than 45.4%.

The EPA's PMF 5.0 model uses the ME-2 program to identify the most optimal factor contributions and profiles. In ME-2, source contribution and composition knowledge as *a priori* information can be included as auxiliary terms of the object function to constrain a model run<sup>21,23</sup>. Three types of constraints are included in the model. There are termed as Ratio, Mass Balance, and Custom. Users can select the Ratio option and use a constant ratio between two different species of a factor to constrain a model run. Users can build an equation by adding one or multiple factor-species on both sides of the equation to constrain a model run under the Mass Balance option. In the Custom option, users can specify a constraint by creating a customized equation to constrain a model run. The customized equation can be built based on either profile (with species as element) or contribution (with sample as element). All the three optional conversations cannot be used to build constraint equations based on the  $^{14}\text{C}$  results as mentioned above, indicating that the  $^{14}\text{C}$  results could not be coupled to EPA PMF 5.0 properly as *a priori* information to constrain a model run and to provide more reliable model results. There are two possibilities: the model software has not provided an appropriate conversation for coupling the  $^{14}\text{C}$  results, or the ME-2 algorithm cannot handle the constraint equations based on the  $^{14}\text{C}$  results. If the ME-2 algorithm can handle the constraint equations, it is vitality and recommended that the model software provides an appropriate conversation.

In order to assess the capacity of the ME-2 algorithm, we performed a preliminary constraint simulation based on BMR-8 to assess the variability of source contributions and availability of  $^{14}\text{C}$  results in the simulation. The preliminary constraint simulation used two additional constraint types (termed Pull Down Maximally and Pull Up Maximally) included in the model. Users can use the two options to pull down maximally and pull up maximally the contributions of one or multiple factors to all species for one or several samples. It should be noted that such a constrained modeling has incapable to provide a more reliable solution. Our concern was whether the range of source contributions from BMR-8 and the constrained modeling based on BMR-8 covers the  $^{14}\text{C}$  results. The coverage indicates that the ME-2 algorithm is able to handle the constraint modeling if the model software provides an appropriate conversation to build constraint equations.

As found above, the largest overestimations were observed in the samples with the highest EC and OC concentrations over the entire sampling period, and the largest overestimations of fossil sources to OC (33.7% and 34.3% without and with consideration of vehicle dust as a fossil source) were observed in the sample with the highest EC concentration. This sample had been collected on January 16, 2012. According to the overestimation, the contribution of two main fossil sources, coal combustion, and vehicle emission on the sampling day were pulled down maximally. Similarly, on June 6, 2012, when the OC concentration reached a maximum during the entire sampling period, biomass burning and sea salt were pulled up maximally. The results of the constrained run showed that Q(robust) and Q(true) increased by 0.51% and 0.63%, respectively, compared to those of BMR-8. Correspondingly, the source profiles and source contributions varied non-significantly compared to those in the base run (SI Figures S26 and S27).

The source contributions to OC and EC modeled by the BMR and the constrained model run were compared to the  $^{14}\text{C}$  results using the two samples with the highest OC and EC concentrations (Fig. 4). Most of the  $^{14}\text{C}$  results fell within the range between those modeled by the base and constrained model runs, suggesting that this model has the potential capacity to provide a reliable source assessment on  $\text{PM}_{2.5}$  in simulations when constrained by powerful *a priori* information, such as  $^{14}\text{C}$  data. However, to exploit this capability, the model must be

able to provide more flexible conversations than those that currently exist. For instance, the sum contributions of several sources to one or several components measured in a sample or some samples collected for a period of time can be defined as a value or range. Once developed, this will provide a powerful tool for using PMF to apportion  $PM_{2.5}$  to sources, because knowledge on the sources of not only carbon-containing matter but also nitrogen- and sulfur-containing matter, in  $PM_{2.5}$  determined by their respective stable isotope measurements can be considered *a priori* information to constrain PMF models<sup>10–12, 24, 25</sup>. Nitrogen-, sulfur-, and carbon-containing matter account for the majority of  $PM_{2.5}$  mass concentrations<sup>18, 26</sup>. Therefore, if these three types of matter are handled as *a priori* information, the resulting constrained PMF models could provide highly reliable source apportionments of  $PM_{2.5}$ , which is vital for designing  $PM_{2.5}$  control strategies.

## Conclusion

The utility of combining PMF with  $^{14}C$  for source apportionment was assessed using  $PM_{2.5}$  data collected for 14 months at a national atmospheric background station in North China. Four model experiments with six to nine factors were run and the most physically interpretable PMF solutions were determined by the comparison of their results. The best solutions were compared to the  $^{14}C$  results of four seasonally averaged samples and three outlier samples with the highest  $PM_{2.5}$ , OC and EC concentrations. The comparison showed that PMF modeling can capture effectively the  $PM_{2.5}$  source patterns with two (8.9% and 1.3%) and three (33.7%, 14.8%, and 2.2%) inappropriate source apportionments for the seasonal and outlier samples. The contribution of sources that could not be classified as either fossil or non-fossil sources in the PMF solution and the errors between the modeled and measured concentrations weakened the validity of the comparison. Based on the two factors, an index was developed for choosing appropriate  $^{14}C$  measurement samples for coupling with PMF simulation. Potential coupling between PMF and  $^{14}C$  was examined by a constrained PMF run using  $^{14}C$  results as prior information. Results show that such a constrained run can obtain a more reliable solution, but PMF should provide a flexible dialog to execute the constraint. For instance, the sum contributions of several sources to one or several components measured in a sample or some samples collected for a period of time can be defined as a value or range. The design can be used to combine PMF models with prior information on nitrogen- and sulfur- containing material in  $PM_{2.5}$  determined by their respective stable isotope measurements, and exhibits the promising prospect for source apportionment.

## Methods

**Site description and sample collection.** Tuoji Island is a small island with an area of 7.1 km<sup>2</sup> located at the demarcation line between the Bohai Sea and the Yellow Sea. It is located 40 km north of the Shandong Peninsula, 300 km east of the Beijing-Tianjin-Hebei region, and 70 km south of the Liaodong Peninsula (SI, Figure S28). There is no industrial activity on the island, and the inhabitants support themselves by fishing. A national station for background atmospheric monitoring is located on the northwestern tip of Tuoji Island (38.188°N, 120.741°E) (SI Figure S29). The sampling platform at this station is about 10 m above ground level<sup>27</sup>.

At the sampling platform,  $PM_{2.5}$  samples were collected at the sampling platform every three days from December 2011 to January 2013. The sampling time started at 10:00 local standard time and sampling continued for 24 h. Samples were collected on quartz fiber filters (QM-A, 20.3 × 25.4 cm<sup>2</sup>, Whatman, heated at 500 °C for 8 h before use) using a high-volume sampler (Hi-Vol 3000; Ecotech, Australia) at a flow rate of 1.13 m<sup>3</sup> min<sup>-1</sup><sup>27</sup>. We selected samples from each month distributed evenly throughout the month and analyzed a total of 70 samples in this study.

**Chemical speciation and  $^{14}C$  measurements.** The mass concentrations of  $PM_{2.5}$  were analyzed gravimetrically using a Sartorius MC5 electronic microbalance. Organic carbon (OC) and elemental carbon (EC) were analyzed with a Desert Research Institute Model 2001 carbon analyzer (Atmoslytic Inc., Calabasas, CA, USA). Water-soluble ions (i.e., sodium [Na<sup>+</sup>], ammonium [NH<sub>4</sub><sup>+</sup>], potassium [K<sup>+</sup>], magnesium [Mg<sup>2+</sup>], calcium [Ca<sup>2+</sup>], chloride [Cl<sup>-</sup>], nitrate [NO<sub>3</sub><sup>-</sup>], and sulfate [SO<sub>4</sub><sup>2-</sup>]) were determined with ion chromatography (ICS-3000; Dionex Ltd., Sunnyvale, CA, USA). Metals (i.e., vanadium [V], manganese [Mn], iron [Fe], chromium [Cr], nickel [Ni], copper [Cu], zinc [Zn], arsenic [As], cadmium [Cd], and lead [Pb]) were measured with inductively coupled plasma mass spectrometry (ELAN DRC II; Perkin Elmer Ltd., Hong Kong). The methods have previously been described<sup>28–30</sup> and details are provided in SI Text S4.

OC was split into the water-soluble organic carbon (WSOC) and water-insoluble organic carbon (WIOC) fractions.  $^{14}C$  was measured in WSOC, WIOC, and EC, respectively<sup>17</sup>. Briefly, WSOC and WIOC were separated with Milli-Q water, along with EC, and each fraction was subsequently converted into carbon dioxide (CO<sub>2</sub>). The CO<sub>2</sub> from the three fractions was cryogenically trapped and reduced to graphite for accelerator mass spectrometry (AMS) target preparation<sup>31–33</sup>. The graphite targets were prepared using the graphitization line at the Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, and the modern carbon fractions ( $f_m$ ) in the graphite samples were determined with a compact accelerator mass spectrometry (AMS, National Electrostatics Corp., Middleton, WI, USA) at Peking University. To take the impact of the nuclear bomb in the 1950s and 1960s into account, the non-fossil carbon fraction ( $f_c$ ) in the samples was defined as  $f_c = f_m/1.10$  for EC and  $f_c = f_m/1.06$  for OC and the fossil carbon fraction ( $f_f$ ) was defined as  $f_f = 1 - f_c$ <sup>34, 35</sup>. This method was previously described in detail<sup>18</sup> and is presented in SI Text S4.

**Source apportionment model and capacity assessment.** We used the Environmental Protection Agency's (EPA's) PMF 5.0 model to explore combining PMF with  $^{14}C$  measurements for the source apportionment of  $PM_{2.5}$ . The model decomposes a matrix into two matrices using multiple iterations of the Multilinear Engine (ME-2)<sup>23</sup>. The optimal solutions for a model run are identified as those with the smallest sums of the squared residuals between the products of the two matrices minus the original matrix divided by the corresponding

uncertainty matrix. During the iteration, two random matrices are generated, and then modified systematically along the direction provided by the conjugate gradient algorithm to find the best-fitting solution<sup>23</sup>. The squared residuals are indicated as Q(true) and Q(robust) in the model. Q(true) is the goodness-of-fit parameter calculated from all of the points. Q(robust) is the goodness-of-fit parameter calculated after excluding points with an uncertainty-scaled residual >4 not fit by the model. The best-fitting solution from a set of random matrices is considered a local minimum, rather than the global minimum, for a decomposed matrix. To maximize the chance of reaching the global minimum, the model should be run many times with different starting points, and the solution with the smallest Q values among the runs is used as the optimal solution and is often called the base model run (BMR). The model's guide suggests that the most optimal solution among 100 model runs can be considered the global minimum for the final source apportionment<sup>36</sup>. The ratio of Q(true)/Q(exp) is an effective parameter for assessing the PMF solution. In the model, the bootstrap enhanced by displacement of factor elements (BS-DISP) method is used to estimate the errors associated with both random and rotational ambiguity. If the change in Q of the BS-DISP estimation is <1% of the Q(robust) from the corresponding BMR, the BMR results are considered to have non-significant rotational ambiguity<sup>36,37</sup>.  $F_{\text{peak}}$  can be used to assess whether the model results fill the solution space by rotating a given solution. If the increase in the Q value due to the  $F_{\text{peak}}$  rotation is >5% of the Q(robust) from the corresponding BMR, the BMR solution can be perceived as covering the full solution spaces<sup>36,37</sup>. Details of these processes are described in the user guide of the model<sup>31</sup>.

In this study, 21 chemical species (OC, EC,  $\text{NO}_3^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{Cl}^-$ ,  $\text{NH}_4^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Na}^+$ , As, Cr, Cd, Cu, Fe, Mn, Ni, Zn, Pb, V, and unrecognized components) were loaded into the model to quantitatively apportion  $\text{PM}_{2.5}$  to sources. The species method detection limits shown in SI Text S4 were used to calculate uncertainties for each sample according to the equation-based uncertainty method. Four model experiments with six to nine factors were each run 100 times. The best solution of each model experiment (i.e., the BMR) was identified as that with the minimum Q(robust) value. The errors associated with both random and rotational ambiguity of the four model experiments were examined from the respective BS-DISP estimations based on the respective BMRs. The solutions of the four BMRs were examined from the  $F_{\text{peak}}$  model run with strengths of -1.0, -0.5, 0.5, 1.0, and 1.5.

To combine PMF with  $^{14}\text{C}$  measurements for the source apportionment of  $\text{PM}_{2.5}$ , we first focused on the PMF model performance by comparing fossil and non-fossil source contributions determined from the  $^{14}\text{C}$  analysis, classified as OC and EC from the PMF results using several specific samples. For the comparison, we used four samples merged seasonally and three outlier samples of  $^{14}\text{C}$  measurements. The seasonally merged samples were made by pooling sample punches of equal sizes cut from samples collected in the first four seasons of the study period. For the outlier samples, we selected the samples with the highest OC, EC, and  $\text{PM}_{2.5}$  concentrations in the entire sampling period. For the comparison, the  $^{14}\text{C}$  fraction of OC was calculated from the concentration-weighted WSOC and WIOC fractions as:

$$f_{\text{OC}} = (f_{\text{WSOC}} \times C_{\text{WSOC}} + f_{\text{WIOC}} \times C_{\text{WIOC}}) / (C_{\text{WSOC}} + C_{\text{WIOC}}), \quad (5)$$

where  $f_{\text{OC}}$ ,  $f_{\text{WSOC}}$ , and  $f_{\text{WIOC}}$  are the non-fossil carbon fractions of OC, WSOC, and WIOC, respectively, and  $C_{\text{WSOC}}$  and  $C_{\text{WIOC}}$  are the WSOC and WIOC concentrations, respectively. The modeled source contributions were classified into two groups according to their fossil and non-fossil carbon sources. The contribution fractions of fossil or non-fossil carbon sources to OC and EC were subsequently compared to the  $^{14}\text{C}$  results of the seven samples. In our previous study<sup>18</sup>, the contribution fractions (R) of non-fossil or fossil sources to OC or EC classified from the PMF results were determined by:

$$R_{ij} = \frac{\sum_{k=1}^n g_{ik} f_{kj}}{\sum_{k=1}^p g_{ik} f_{kj}}, \quad (6)$$

where  $g$  and  $f$  are the factor contributions and factor profiles, respectively,  $i$  represents the OC or EC species,  $j$  is a specified sample,  $n$  is the number of fossil or non-fossil carbon sources, and  $p$  is the number of all of the sources. The calculation is based on the total OC or EC concentrations based on the PMF model rather than the measured concentrations, indicating that the errors between the modeled and measured concentrations are ignored in these comparisons<sup>18</sup>. For a more reasonable comparison and assessment, in this study we replaced the modeled total concentration in Equation (6) with the corresponding measured concentration:

$$R_{ij} = \frac{\sum_{k=1}^n g_{ik} f_{kj}}{X_{ij}}, \quad (7)$$

where  $X_{ij}$  is the measured OC or EC concentration and the other parameters are the same as those in Equation (6). Furthermore, we used backward trajectories (<http://www.arl.noaa.gov/ready.html>)<sup>38</sup> and fire counts (<https://firms.modaps.eosdis.nasa.gov/firemap/>) to assess the potential sources of  $\text{PM}_{2.5}$ . A distance analysis and Pearson correlation were used to quickly identify the source types and their contributions to  $\text{PM}_{2.5}$ . Detailed methods are described in SI Text 5.

## References

- U. S. EPA. Final report: Integrated science assessment for particulate matter. Report No. EPA/600/R-08/139F, 1–983 (U.S. Environmental Protection Agency, Washington, DC, 2009).
- Balachandran, S. *et al.* Bayesian-based ensemble source apportionment of  $\text{PM}_{2.5}$ . *Environ. Sci. Technol.* **47**, 13511–13518 (2013).
- Belis, C. A., Karagulian, E., Larsen, B. R. & Hopke, P. K. Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe. *Atmos. Environ.* **69**, 94–108 (2013).
- Pant, P. & Harrison, R. M. Critical review of receptor modelling for particulate matter: A case study of India. *Atmos. Environ.* **49**, 1–12 (2012).



5. Zheng, M. *et al.* Review of PM<sub>2.5</sub> source apportionment methods in China. *Acta Sci. Nat. Univ. Pekinensis* **50**, 1141–1154 (2014).
6. Song, Y. *et al.* Source apportionment of PM<sub>2.5</sub> in Beijing using principal component analysis/absolute principal component scores and UNMIX. *Sci. Total Environ.* **372**, 278–286 (2006).
7. Bove, M. C. *et al.* An integrated PM<sub>2.5</sub> source apportionment study: Positive Matrix Factorisation vs. the chemical transport model CAMx. *Atmos. Environ.* **94**, 274–286 (2014).
8. Xie, M. *et al.* Positive matrix factorization of PM<sub>2.5</sub>: Comparison and implications of using different speciation data sets. *Environ. Sci. Technol.* **46**, 11962–11970 (2012).
9. Tao, J. *et al.* Uncertainty assessment of source attribution of PM<sub>2.5</sub> and its water-soluble organic carbon content using different biomass burning tracers in positive matrix factorization analysis — A case study in Beijing, China. *Sci. Total Environ.* **543**(Part A), 326–335 (2016).
10. Han, X. *et al.* Using stable isotopes to trace sources and formation processes of sulfate aerosols from Beijing, China. *Sci. Rep.* **6**, 29958 (2016).
11. Pan, Y. *et al.* Fossil fuel combustion-related emissions dominate atmospheric ammonia sources during severe haze episodes: Evidence from <sup>15</sup>N-stable isotope in size-resolved aerosol ammonium. *Environ. Sci. Technol.* **50**, 8049–8056 (2016).
12. Zheng, M. *et al.* Spatial distribution of carbonaceous aerosol in the southeastern United States using molecular markers and carbon isotope data. *J. Geophys. Res. Atmos.* **111**, D10S06 (2006).
13. Szidat, S. Sources of Asian haze. *Science* **323**, 470–471 (2009).
14. Szidat, S. *et al.* Radiocarbon (<sup>14</sup>C)-deduced biogenic and anthropogenic contributions to organic carbon (OC) of urban aerosols from Zürich, Switzerland. *Atmos. Environ.* **38**, 4035–4044 (2004).
15. Zhang, Y. L. *et al.* Fossil vs. non-fossil sources of fine carbonaceous aerosols in four Chinese cities during the extreme winter haze episode of 2013. *Atmos. Chem. Phys.* **15**, 1299–1312 (2015).
16. Liu, D. *et al.* The use of levoglucosan and radiocarbon for source apportionment of PM<sub>2.5</sub> carbonaceous aerosols at a background site in East China. *Environ. Sci. Technol.* **47**, 10454–10461 (2013).
17. Liu, J. *et al.* Source apportionment using radiocarbon and organic tracers for PM<sub>2.5</sub> carbonaceous aerosols in Guangzhou, South China: Contrasting local- and regional-scale haze events. *Environ. Sci. Technol.* **48**, (12002–12011) (2014).
18. Zong, Z. *et al.* Source apportionment of PM<sub>2.5</sub> at a regional background site in North China using PMF linked with radiocarbon analysis: insight into the contribution of biomass burning. *Atmos. Chem. Phys.* **16**, 11249–11265 (2016).
19. Zhang, Y. *et al.* High contribution of nonfossil sources to submicrometer organic aerosols in Beijing, China. *Environ. Sci. Technol.* doi:10.1021/acs.est.1027b01517 (2017).
20. Kessler, R. Prevention: Air of danger. *Nature* **509**, S62–S63 (2014).
21. Paatero, P. & Hopke, P. K. Rotational tools for factor analytic models. *J. Chemom.* **23**, 91–100 (2009).
22. Nguyen, Q. T. *et al.* Source apportionment of particles at Station Nord, North East Greenland during 2008–2010 using COPREM and PMF analysis. *Atmos. Chem. Phys.* **13**, 35–49 (2013).
23. Paatero, P. The multilinear engine—A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J. Comput. Geog. Sci.* **8**, 854–888 (1999).
24. Zong, Z. *et al.* First assessment of NO<sub>x</sub> sources at a regional background site in North China using isotopic analysis linked with modeling. *Environ. Sci. Technol.* **51**, 5923–5931 (2017).
25. Wang, Y.-L. *et al.* Source appointment of nitrogen in PM<sub>2.5</sub> based on bulk δ<sup>15</sup>N signatures and a Bayesian isotope mixing model. *Tellus Ser. B: Chem. Phys. Meteorol.* **69**, 1299672 (2017).
26. Huang, R.-J. *et al.* High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* **514**, 218–222 (2014).
27. Wang, X. *et al.* Impact of agricultural waste burning in the Shandong Peninsula on carbonaceous aerosols in the Bohai Rim, China. *Sci. Total Environ.* **481**, 311–316 (2014).
28. Chow, J. C. *et al.* The IMPROVE\_A temperature protocol for thermal/optical carbon analysis: Maintaining consistency with a long-term database. *J. Air Waste Manage. Assoc.* **57**, 1014–1023 (2007).
29. Shahsavani, A. *et al.* Characterization of ionic composition of TSP and PM<sub>10</sub> during the Middle Eastern Dust (MED) storms in Ahvaz, Iran. *Environ. Monit. Assess.* **184**, 6683–6692 (2012).
30. Wang, X., Bi, X., Sheng, G. & Fu, J. Hospital indoor PM<sub>10</sub>/PM<sub>2.5</sub> and associated trace elements in Guangzhou, China. *Sci. Total Environ.* **366**, 124–135 (2006).
31. Xu, X. *et al.* Modifying a sealed tube zinc reduction method for preparation of AMS graphite targets: Reducing background and attaining high precision. *Nucl. Instrum. Methods Phys. Res. Sect. B* **259**, 320–329 (2007).
32. Zhang, Y. L. *et al.* Development of a preparation system for the radiocarbon analysis of organic carbon in carbonaceous aerosols in China. *Nucl. Instrum. Methods Phys. Res. Sect. B* **268**, 2831–2834 (2010).
33. Wacker, L. *et al.* A versatile gas interface for routine radiocarbon analysis with a gas ion source. *Nucl. Instrum. Methods Phys. Res. Sect. B* **294**, 315–319 (2013).
34. Zong, Z. *et al.* Radiocarbon-based impact assessment of open biomass burning on regional carbonaceous aerosols in North China. *Sci. Total Environ.* **518–519**, 1–7 (2015).
35. Zhang, Y.-L. *et al.* Radiocarbon-based source apportionment of carbonaceous aerosols at a regional background site on Hainan Island, South China. *Environ. Sci. Technol.* **48**, 2651–2659 (2014).
36. Norris, G. & Duvall, R. EPA positive matrix factorization (PMF) 5.0 fundamentals and user guide. Report No. EPA/600/R-14/108, 1-124 (U.S. Environmental Protection Agency National Exposure Research Laboratory, Washington, 2014).
37. Paatero, P., Eberly, S., Brown, S. G. & Norris, G. A. Methods for estimating uncertainty in factor analytic solutions. *Atmos. Meas. Tech.* **7**, 781–797 (2014).
38. Stein, A. F. *et al.* NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. *Bull. Am. Meteorol. Soc.* **96**, 2059–2077 (2015).

## Acknowledgements

This work was financially supported by Key Laboratory of Coastal Environmental Processes and Ecological Remediation of the Chinese Academy of Sciences (CAS) (Grant No: 2016KFJ01), the Natural Scientific Foundation of China (NSFC; Grant Nos: 41471413 and 41473101), Guangzhou Science and Technology Plan Project (No. 201504010002) and the Strategic Priority Research Program of the CAS (Grant No. XDA11020402). The authors gratefully acknowledge the National Oceanic and Atmospheric Administration's Air Resources Laboratory for providing the HYSPLIT transport model and the READY website (<http://www.arl.noaa.gov/ready.html>), and the NASA/GSFC/Earth Science Data and Information System (ESDIS) with funding provided by NASA/HQ for providing the data and imagery from LANCE FIRMS (<https://firms.modaps.eosdis.nasa.gov/firemap/>). This is a contribution of GIGCAS No.IS-2422.

### Author Contributions

The work presented here was carried out in collaboration between all authors. C.T. and X.W. conceived and designed the work, analyzed the data, interpreted the results, and drafted the manuscript; X.W. and Z.Z. performed the experiments; Y.C., C.L., J.L., G.Z., and Y.L. participated in the acquisition, analysis, interpretation of data or provided constructive discussions.

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-10762-8](https://doi.org/10.1038/s41598-017-10762-8)

**Competing Interests:** The authors declare that they have no competing interests.

**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 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017