Response to Reviewer #1's Comments

Anonymous Referee #1:

Major comments:

This study investigated the impacts of anthropogenic and natural emission sources on atmospheric mercury before, during, and after the COVID-19 lockdown based on data measured at a rural site in eastern China. Correlation analysis, an explainable machine learning model, and the PMF model were applied to quantify the impacts of the key factors reflecting anthropogenic and natural sources. The manuscript tried to depict the change of atmospheric mercury behavior caused by the COVID-19 lockdown in China. However, my major concern is that the authors seemed to be a bit arbitrary in drawing conclusions. More solid evidences are required. Moreover, the reliability of the machine learning model and the PMF model needs more rigorous illustration. The novelty of this study also needs to be better addressed. The compensation effect of natural mercury emission when the GEM concentration is reduced has been reported before, and the results from this study could not confirm this effect. In addition, the discussion part in the manuscript needs significant improvement. Therefore, in my opinion, this manuscript is not acceptable for publication on Atmospheric Chemistry and Physics in its current version.

We sincerely thank for the reviewer's in-depth comments and helpful suggestions on this manuscript. Based on the specific comments, we have responded to all the comments point-by-point and made corresponding changes in the manuscript as highlighted in red color. The reviewer has raised a number of issues and we quite agree. The manuscript has been significantly revised. We feel the substantial revisions based on the reviewer's comments have greatly improved the quality of this manuscript. Please check the detailed responses to all the comments as below.

Here are some specific comments:

1. Lines 20–23: The conclusion that the decrease of GEM was not as significant as other air pollutants is not convincing.

Response: Thanks for the comment. In order to make this conclusion more solid, the table below shows the changes in GEM and the other measured air pollutants (including SO_2, NO_2, CO, O_3, NO, NO_3^-, SO_4^{2-}, NH_4^+, Ca^{2+}, OC, EC, Pb, Fe, Cr, Se, Ca, Mn, As, Ni, Zn, and BC) before and during the lockdown. It is found that except for SO_2, SO_4^{2-}, and OC, the declines of other air pollutants are all greater than that of GEM. The average decline percentage of those air pollutants is 55%, while that of GEM is 26%. To make the expression more accurate, we change the sentence as "At a regional site in Eastern China, an intensive measurement was performed, showing obvious decreases of gaseous elemental mercury (GEM) during the COVID-19 lockdown, while not as significant as most of the other measured air pollutants" in **Line 23** of the revision.

As for the original writing in the first paragraph in Section 3.2, "Figure S4 further shows the reduction rates of GEM, SO_2, NO_2, CO, EC, Pb, As, and BC during the lockdown were 26%, 9%, 56%, 33%, 38%, 36%, 34%, and 51%, respectively, compared to pre-lock. Except for SO2, GEM presented lower reduction rate than the other air pollutants, probably suggesting the different release mechanism of GEM from the other air pollutants.", it is changed as "Table S2 further shows the reduction rates of gaseous pollutants SO2, NO2, NO, and CO were 9%, 56%, 64%, and 33%, respectively, compared to those before the lockdown. While O_3 showed almost one-fold increase due to the strongly depressed titration effect from substantial reduced NOx emissions during the lockdown (Huang et al., 2021; Yang et al., 2021). As for the primary trace elements such as Pb, Fe, Cr, Se, Ca, Mn, As, Ni, and Zn, their reduction rates ranged from 34% - 73%. As for the main chemical components in $PM_{2.5}$, NO_3^-, NH_4^+, and BC were strongly reduced by 58%, 45%, and 51%, while SO_4^{2-} and OC were less reduced by 20% and 16%, respectively. Except for SO_2, SO_4^{2-}, and OC, GEM presented lower reduction rate than the other air pollutants, probably indicating the discrepancy in key sources for different air pollutants." in **Line 276-284** of the revision.

Table. The changes of GEM and the other measured air pollutants before and during the lockdown.

	before	during	relative change (%)
GEM	2.78	2.06	26
SO_2	7.47	6.76	9
NO_2	53.52	23.76	56
CO	0.91	0.61	33
O_3	42.13	83.10	-97
NO	11.53	4.17	64
NO_3^-	23.37	9.73	58
SO_4^{2-}	9.97	7.99	20
NH_4^+	10.88	5.98	45
Ca_2^+	0.14	0.03	77
OC	5.32	4.47	16
EC	2.45	1.52	38
Pb	32.62	21.02	36
Fe	372.22	137.59	63
Cr	5.82	1.06	82
Se	3.37	2.28	32
Ca	67.15	19.86	70
Mn	36.63	9.92	73
As	8.09	5.36	34
Ni	4.02	1.82	55
Zn	112.76	35.26	69
BC	1.87	0.93	51

Note: the units of GEM and all trace elements are ng/m^3. The units of the other air pollutants are μg/m^3.

2. Lines 46–58: It is not quite appropriate to use the past tense in these sentences. More updated literatures could be used, e.g., Streets et al. (2019) and Steenhuisen and Wilson (2019) for global anthropogenic Hg emissions, Liu et al. (2019) for anthropogenic Hg emissions in China, and Pirrone et al. (2010) and Outridge et al. (2018) for global natural Hg emissions.

Response: Thanks for the suggestion, we changed the past tense to the present tense and updated the literatures according to reviewer's suggestions in **Line 46-55** of the revision as follows. "Mercury in the atmosphere derives from both anthropogenic emissions and natural processes. The main anthropogenic sources of atmospheric mercury include coal combustion, nonferrous smelters, cement production, waste incineration, and mining (Wu et al., 2018; Wu et al., 2016). The amount of mercury in the atmosphere directly emitted by anthropogenic activities accounted for about 30% of global mercury emissions (Streets et al., 2019; Steenhuisen and Wilson, 2019) and China is the country with the largest anthropogenic atmospheric mercury emissions in the world (Liu et al., 2019). The natural sources of mercury in the atmosphere are mainly from the exchange processes between natural surfaces (e.g., soil, vegetation, and water) and the atmosphere (Outridge et al., 2018; Pirrone et al., 2010). Unlike anthropogenic emissions, natural releases of mercury are passive emissions and are susceptible to various environmental factors, such as meteorological parameters…".

3. Section 2.2: The size of the input dataset should be given. The results from model verification should be introduced in detail.

Response: Thanks for the suggestion. We add the size of the input dataset in the revision, the relevant sentence has been changed as "The long-term observational GEM (hourly data from March 1, 2015 to February 28, 2019; n = 17532) in Shanghai was the target variable for training, and the corresponding air pollutants (SO_2, CO, O_3, NO_2, and $PM_{2.5}$) and meteorological data (air temperature, relative humidity, and wind speed) were chosen as input variables for training." in **Line 175-179** of the revision.

We add details about the model verification in **Line 182-193** of the revision. "To verify the accuracy of the trained neural network model, we compared the observed (not included in the training data set) and simulated GEM concentrations of DSL from January 1 to February 26, 2020, and found that they exhibited a reasonably good correlation with the correlation coefficient (R^2) of 0.67. To test the applicability of the model on the regional scale, we compared the observed and simulated GEM concentrations in Suzhou, Ningbo, Nanjing, and Hefei (Figure S2). In Nanjing and Suzhou, the observed and simulated daily GEM showed consistence with R^2 values of 0.52 and 0.71, respectively. In Ningbo, the observed and simulated GEM in summer and winter also showed consistence with R^2 values of 0.64 and 0.65, respectively. A low bias was derived between the observed and simulated seasonal GEM in

Hefei. This suggested that it was feasible to use the trained ANN model to simulated the GEM concentrations in Shanghai and even the Yangtze River Delta region."

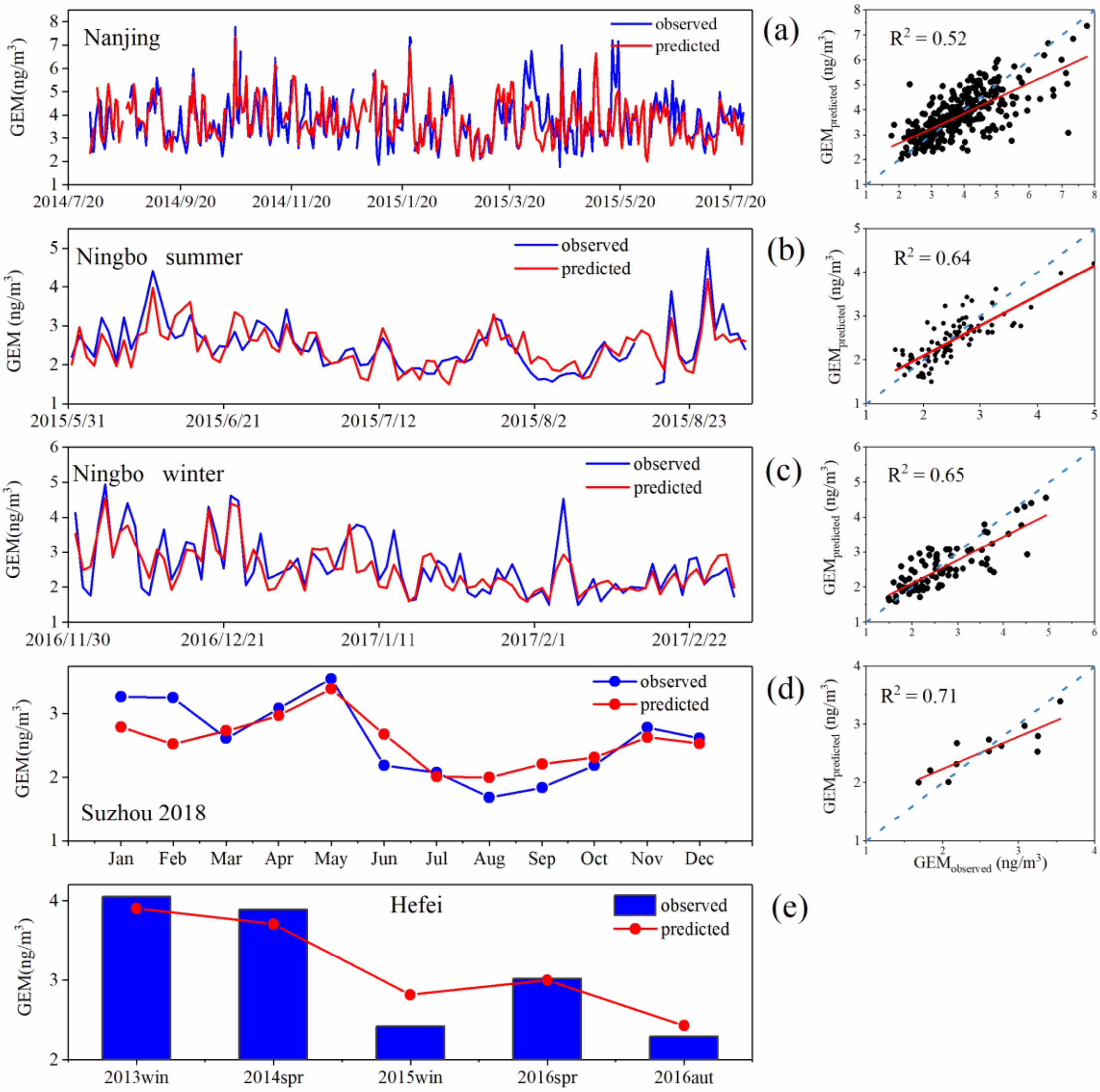

Figure S2. Comparison of observed (blue line) and simulated (red line) GEM concentrations in (a) Nanjing, (b) summer Ningbo, (c) winter Ningbo, (d) Suzhou, and (e) Hefei. The left panel is their time series, and the right panel is their corresponding correlation. The blue dotted line represents the y = x reference line.

4. Section 2.4: Which factors were considered in the PMF model? How were they determined?

Response: In this study, we explored the number of factors being from three to eight, with the optimal solutions determined by the slope of the Q value versus the number of factors. The Q value is the sum of the square of the difference between the measured and modeled concentrations weighted by the concentration uncertainties, and needs to be minimized before the PMF modeled determines the optimal nonnegative factor profiles and contributions (Cheng et al., 2015).

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{X_{ij} - \sum_{k=1}^{p} A_{ik} F_{kj}}{S_{ij}} \right)^2$$

Where X_{ij} represents the concentration of the *j*th contamination in the *i*th sample, m is the total number of the pollutants, and n is the total number of samples. A_{ik} represents the contribution of the kth factor on the ith sample, and F_{kj} represents the mass fraction of the j pollutant in the kth factor. S_{ij} is the uncertainty in the *j*th pollutant on the *i*th factor, and P is the number of factors. For each run in this study, the stability and reliability of the outputs were assessed by referring to the Q value, residual analysis, and correlation coefficients between observed and predicted concentrations. Finally, we found that a six-factor solution showed the most stable results and gave the most reasonable interpretation.

The above writings are added in **Line 224-237** of the revision.

5. Lines 199–200: The case after the "lockdown" could not be called a "rebound". There were just two small GEM pollution episodes. They could also occur during the "lockdown". In fact, the lockdown continued for about six months. It was just that the extent of lockdown was gradually weakening.

Response: We agree with the reviewer that the word "rebound" is inappropriate here, so we change the sentence as "After the lockdown, GEM concentration was slightly higher than that of during the lockdown." in **Line 247-248** of the revision.

6. Lines 230–231: This statement is a bit arbitrary. The lower reduction rate of GEM could be due to the discrepancy in key sources for different air pollutants.

Response: We agree with the reviewer that the lower reduction rate of GEM could be due to the discrepancy in key sources for different air pollutants. In fact, this is exactly what we meant to express here. More discussions about the reduction of GEM compared to the other air pollutants have been detailed in the response to Comment #1. This sentence is revised as "Except for SO_2, SO_4^{2-}, and OC, GEM presented lower reduction rate than the other air pollutants, probably indicating the discrepancy in key sources for different air pollutants." in **Line 283-284** of the revision.

7. Lines 233–237: This statement is too vague. What kind of fossil fuel combustion? Coal combustion or vehicle emissions? Is biomass burning an important source in Shanghai?

Response: Thanks for pointing this issue out. As indicated by previous studies (Qin et al., 2019), source apportionment results indicated vehicle emission was a negligible contributor to mercury, while coal combustion contributed significantly in Shanghai. As for biomass burning, it was not important in Shanghai. However, the straw burning in the adjacent areas of Shanghai was ubiquitous as the Yangtze River Delta is a base for crop cultivation (Yang and Zhao, 2019). To make the expression more accurate, we change the sentence as "This suggested that the main anthropogenic sources of GEM might be coal combustion and biomass burning in Shanghai, which was consistent with the previous studies in the Yangtze River Delta (Qin et al., 2019; Tang et al., 2018)." in **Line 289** of the revision.

8. Lines 238–239: It is not appropriate to regard BC as a proxy for anthropogenic emissions. What kind of anthropogenic emissions?

Response: Sorry for the confusing description, what we wanted to express is that BC can be a proxy for the main anthropogenic sources of GEM. Because BC mainly come from fossil fuels combustion and biomass burning (Briggs and Long, 2016), which is also the main anthropogenic sources of GEM (Streets et al., 2019). We realize that it is not sufficient to use BC as the proxy of main anthropogenic sources of GEM only, so in the revised version, we also add CO and EC as indicators. The sentence is modified as "BC, EC, and CO are mainly from fossil fuels combustion and biomass burning, and can be used as indicators of the main anthropogenic sources of GEM." in **Line 291-292** of the revision.

9. Lines 242–243: Using only the R value or the ratio of GEM/BC to indicate the contribution of anthropogenic sources is not robust. Also, the ratio of GEM/BC should be compared with previous studies.

Response: We agree with the reviewer that it is not robust to indicate the change of GEM sources only through the change of the relationship between GEM and BC in different control stages. In the revised version, in addition to BC, we also discussed the relationship between GEM and CO, EC, so that we can more clearly show the changes in the sources of GEM. As shown in the figure below, the R values between GEM and BC, CO, EC before the lockdown were higher than those during and after the lockdown, which indicates that the influence of anthropogenic sources on GEM is indeed weakened during and after the lockdown by using different parameters. The related sentences have also been changed accordingly in the revision as follows. "As shown in Figure 2, R between GEM and BC, GEM and CO, GEM and EC during and after the lockdown were lower than that before the lockdown", "Different from BC, CO, and EC that are overwhelmingly derived from anthropogenic sources, natural sources such as surface emission and ocean release also contribute significantly to GEM (Obrist et al., 2018). Hence, the ratio of GEM/BC, GEM/CO, and GEM/EC can be simply applied as indicators to reveal the relative importance of anthropogenic versus natural sources." in **Line 296-300** of the revision.

Few studies have addressed the ratio of GEM/BC, while the GEM/CO ratio has been used to analyze the source of GEM in many studies. In this study, the GEM/CO ratio during the lockdown period was 4.0×10^{-6}, which was significantly higher than the anthropogenic GEM/CO emission ratio in mainland China, South Asia, and Indochinese Peninsula, whose values were 2.7, 2.6, and 1.5×10^{-6}, respectively (Fu et al., 2015), also higher than the GEM/CO ratio observed in Nanjing (3.1×10^{-6}) and Beijing (1.5×10^{-6}) in winter (Zhang et al., 2013; Zhu et al., 2012). This implies that during the lockdown, the impact of anthropogenic sources on GEM is weakened and natural sources become more important. We added relevant sentences

"The GEM/CO ratio has been used to analyze the sources of GEM in many studies. In this study, the GEM/CO ratio during the lockdown period was 4.0×10^{-6}, which was significantly higher than the anthropogenic GEM/CO emission ratio in mainland China, South Asia, and Indochinese Peninsula, whose values were 2.7, 2.6, and 1.5×10^{-6}, respectively (Fu et al., 2015), also higher than the GEM/CO ratio observed in Nanjing (3.1×10^{-6}) and Beijing (1.5×10^{-6}) in winter (Zhang et al., 2013; Zhu et al., 2012)." in **Line 304-309** of the revision.

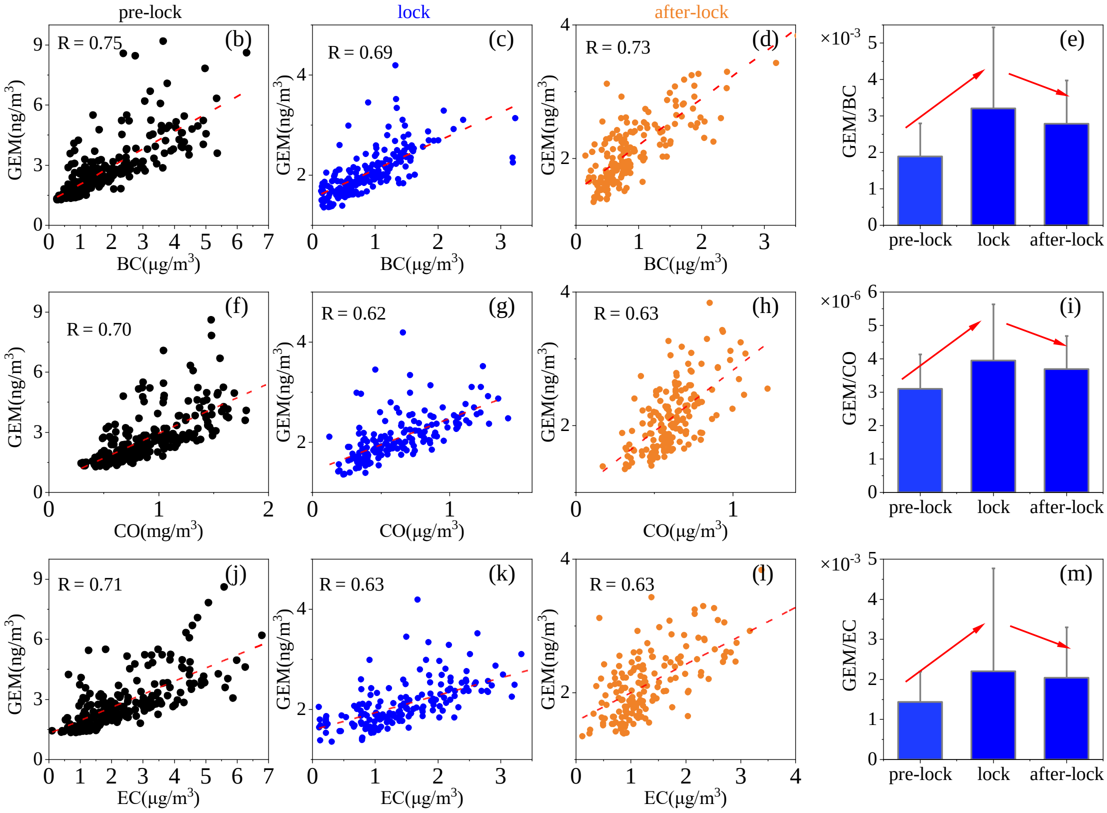

Figure. Relationship between GEM and BC, CO, EC and the change of their ratios before, during, and after the lockdown.

10. Lines 262–266: The increase of the R value didn't necessarily imply the enhanced role of natural sources. It could be that other meteorological factors took the lead in affecting GEM concentration before and during the lockdown.

Response: Thanks for the comment. As can be seen from the table below, the main meteorological factors (including temperature, wind speed, relative humidity, planetary boundary layer height and air pressure) before and during the lockdown did not change significantly. And the 3-days backward trajectory cluster analysis as shown in the Figure below indicated that the transport patterns differed little between these two periods. Therefore, we believe that the reason for the increase in R value between GEM and temperature should not be caused by changes in meteorological factors. Temperature, as an important factor affecting the natural releases of GEM (Pannu et al., 2014; Zhang et al., 2021), has a significantly better correlation with GEM during the lockdown than that before the lockdown, the most likely reason is that the effect of natural release on GEM has become more important, especially as the average temperature during the lockdown is lower than that before the lockdown. In order to make the statement more rigorous, we change the sentence as "This might indicate the enhanced role of natural sources on GEM concentration due to the lockdown control measures." in **Line 324-325** of the revision.

Table. Changes in major meteorological factors before, during, and after the lockdown.

	Temperature (°C)	WS(m/s)	RH (%)	PBL(m)	Pressure (Pa)
before	7.2	1.8	83.1	416.6	1024.9
during	6.1	1.9	77.7	483.0	1024.1
after	9.9	2.3	74.4	439.8	1024.5

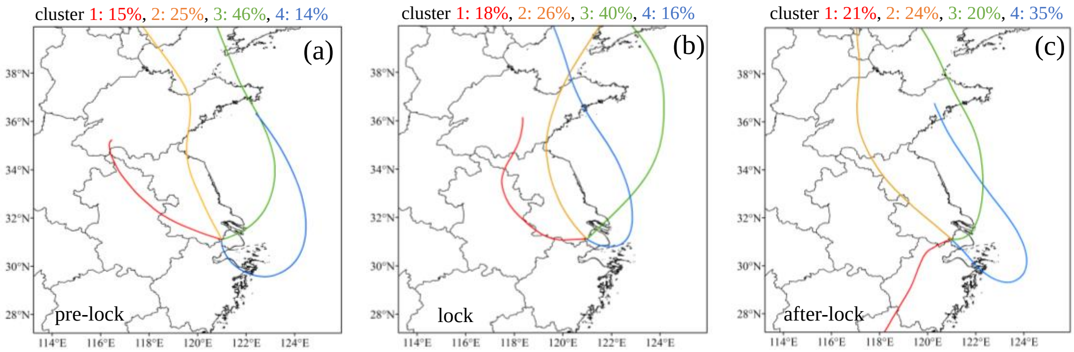

Figure. HYSPLIT 3-days backward trajectory cluster analysis at DSL (a) before, (b) during, and (c) after the lockdown.

11. Lines 266–277: It could also be that different meteorological factors dominate the influence on GEM concentration before, during and after the lockdown.

Response: Thanks for the comment. As responded above, the table below shows the main meteorological factors (including temperature, wind speed, relative humidity, planetary boundary layer height and air pressure) before and during the lockdown did not change significantly. And the 3-days backward trajectory cluster analysis as shown in the Figure below indicated that the transport patterns differed little between these two periods. Therefore, we suggested that it is unlikely that changes in meteorological conditions dominated the changes in GEM concentration before, during and after the lockdown.

Table. Changes in major meteorological factors before, during, and after the lockdown.

	Temp(°C)	WS(m/s)	RH (%)	PBL(m)	Pressure (Pa)
before	7.2	1.8	83.1	416.6	1024.9
during	6.1	1.9	77.7	483.0	1024.1

after	9.9	2.3	74.4	439.8	1024.5

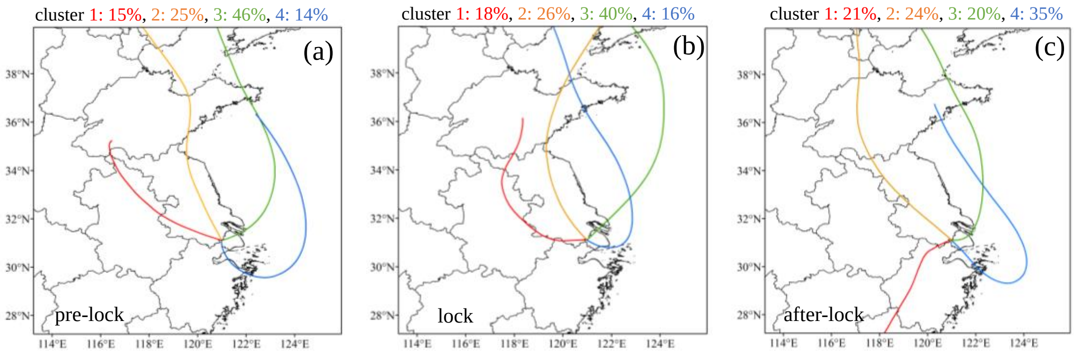

Figure. HYSPLIT 3-days backward trajectory cluster analysis at DSL (a) before, (b) during, and (c) after the lockdown.

12. Line 278: These were very weak evidences. The word "confirmed" is too strong.

Response: Thanks for pointing this out. We changed the word "confirmed" to "possibly suggested" in **Line 338** of the revision.

13. Section 3.3: There should be a training dataset and a test dataset to check if the model is overfitting.

Response: Thanks for the suggestion. In this study, the training dataset includes the target variable GEM (hourly data, n = 17532) observed in Shanghai from March 1, 2015 to February 28, 2019, and the input variables were corresponding air pollutant data (SO_2, CO, O_3, NO_2, and $PM_{2.5}$) and meteorological data (air temperature, relative humidity, and wind speed). During the training of the model, the data set were randomly divided into three parts, i.e., 70% for training, 15% for validation, and 15% for testing. In order to verify the feasible of the trained model, we first compared the observed and simulated GEM in Shanghai from January 1 to February 26, 2020 (not included in the model training data set), and found that they are well correlated (as shown in the Figure below), which indicated that it is feasible to use the model to simulate the GEM concentration in Shanghai.

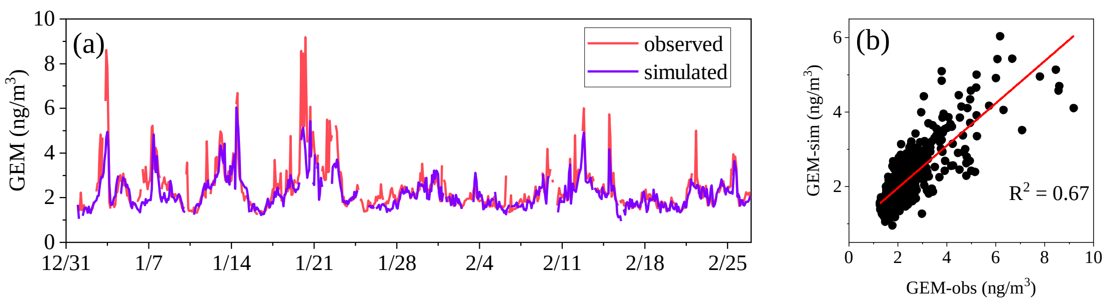

Figure. Time-series of observed and ANN-simulated GEM concentrations during the study period.

To test the applicability of the model on a regional scale, we compared the observed and simulated GEM concentrations in Suzhou, Ningbo, Nanjing, and Hefei (Figure below). In Nanjing and Suzhou, the observed and simulated daily GEM shows consistence with R^2 values of 0.52 and 0.71, respectively. In Ningbo, the observed and simulated GEM in summer and winter also show consistence with R^2 values of 0.64 and 0.65, respectively. A low bias was derived between the observed and simulated seasonal GEM in Hefei. This suggested that it is feasible to use the trained ANN model to simulated the GEM concentrations in Shanghai and even the Yangtze River Delta region.

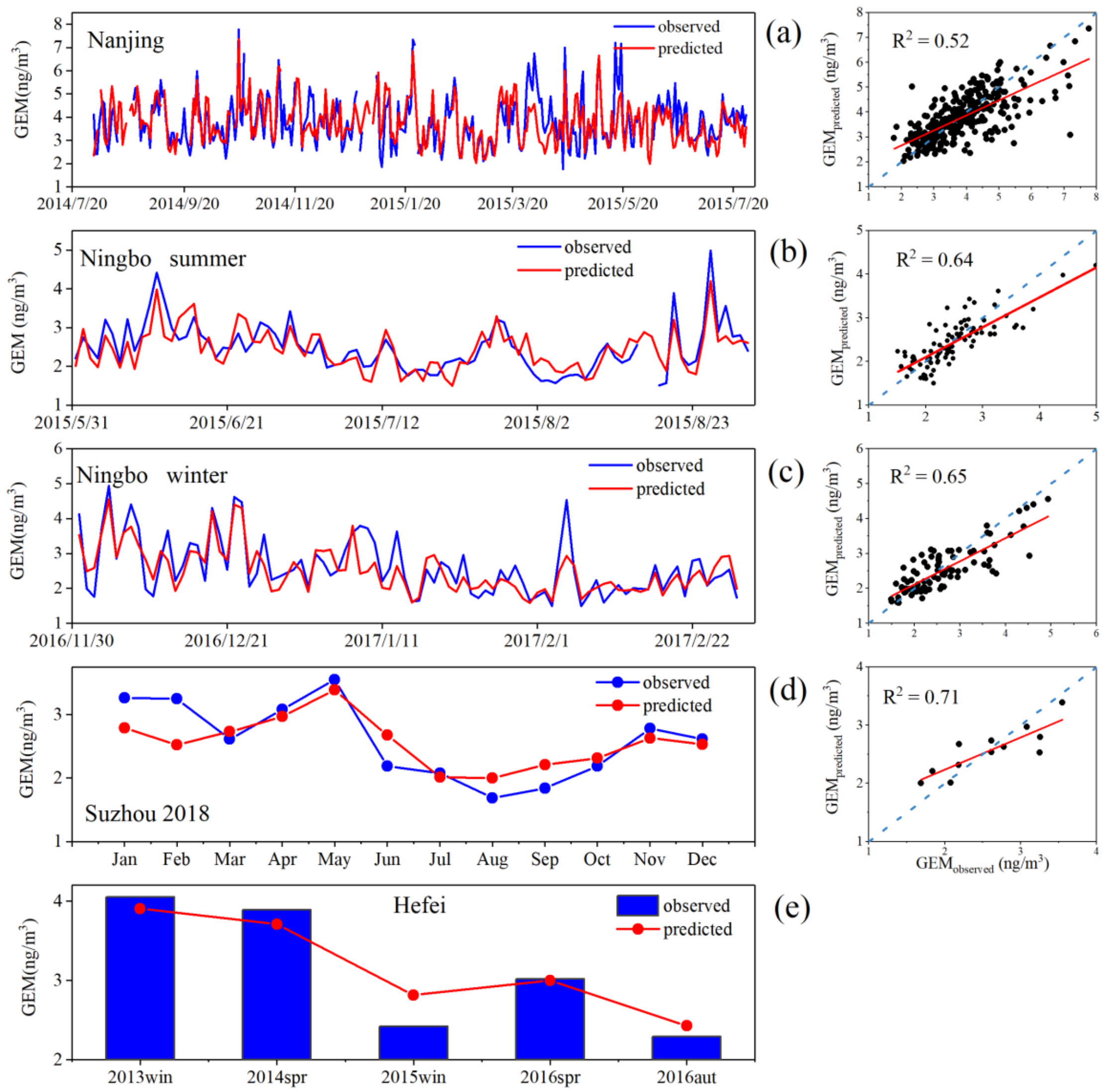

Figure. Comparison of observed (blue line) and simulated (red line) GEM concentrations in (a) Nanjing, (b) summer Ningbo, (c) winter Ningbo, (d) Suzhou, and (e) Hefei. The left panel is their time series, and the right panel is their corresponding correlation. The blue dotted line represents the y = x reference line.

We added detail description “To verify the accuracy of the trained neural network model, we compared the observed (not included in the training data set) and simulated GEM concentrations of DSL from January 1 to February 26, 2020, and found that they have a good correlation, the correlation coefficient (R^2) is 0.67. To test the applicability of the model on a regional scale, we compared the observed and simulated GEM concentrations in Suzhou, Ningbo, Nanjing, and Hefei (Figure S2). In Nanjing and Suzhou, the observed and simulated daily GEM showed consistence with R^2 values of 0.52 and 0.71, respectively. In Ningbo, the observed and simulated GEM in summer and winter also showed consistence with R^2 values of 0.64 and 0.65, respectively. A low bias was derived between the observed and simulated seasonal GEM in Hefei. This suggested that it is feasible to use the trained ANN model to simulated the GEM concentrations in Shanghai and even the Yangtze River Delta region." in **Line 182-193** of the revision.

14. Figure 4(d-k): The influencing patterns of each factor should be discussed.

Response: Thanks for the suggestion. We added a detailed description of the influencing patterns of each factor in **Line 368-381** of the revision.

"As shown in Figure 4d-f, with the increase of $PM_{2.5}$, CO, and SO_2 concentrations, their corresponding SHAP values increased accordingly. Previous studies have shown that GEM, $PM_{2.5}$, CO, and SO_2 shared common anthropogenic sources such as the combustion of fossil fuels and biomass (Chong et al., 2019; Fu et al., 2015), thus interpreting the positive effect of various anthropogenic emission sources on GEM. Similar relationship was also found for temperature and relative humidity with their corresponding SHAP values (Figure 4g-h). Since temperature and relative humidity are important factors affecting the natural release of GEM from natural surfaces (Pannu et al., 2014; Wang et al., 2016), the positive influence of natural surface emissions on GEM was expected. In contrast, the SHAP value of wind speed negatively correlated with the magnitude of wind speed (Figure 4i), indicating the diffusion/accumulation effect of wind speed on GEM. The SHAP values of NO_2 and O_3 did not show obvious correlations with their concentrations (Figure 4j-k). One of the main sources of NO_2 was vehicle emission, which contributed little to GEM. As for O_3, its oxidation on GEM was also limited. Thus, neither NO_2 or O_3 exhibited considerable effects on regulating the GEM variation."

15. Lines 334–338: Does the SHAP value reflect the absolute impact of the factor or the relative contribution of it? If latter, the statement is incorrect. If former, the statement is not necessarily correct either.

Response: Thanks for the comments. According to the algorithm of SHAP value, the SHAP value of a feature for a query point explains the deviation of the prediction for the query point from the average prediction due to the feature. For each query point, the sum of the SHAP values for all features corresponds to the total deviation of the prediction from the average. In this study, the SHAP value of a feature at a specific time represents its influence on the deviation of the GEM concentration at that time from the average GEM concentration throughout the whole study period. For example, the SHAP value of temperature can be used to represent the influence of natural sources on GEM concentration at corresponding time deviating from the average GEM concentration throughout the whole study period, and the SHAP values of $PM_{2.5}$ and CO can be used to represent the influence of major anthropogenic sources. Thus, the SHAP value reflects the absolute impact of the factor but not the relative contribution.

As for the statement in Line 334–338, we agree with the reviewer that the conclusion which was based on the SHAP value analysis needs more evidence. Based on the negative correlation between the SHAP value of temperature and $PM_{2.5}$, CO, we suggested a possible mechanism that the increase of anthropogenic emissions of GEM might inhibit the natural release of GEM to some extents. However, this conclusion should be confirmed and thus the PMF modeling was applied in Section 3.4. As shown in Figure 6b, the absolute contribution of natural surface emissions to GEM during the lockdown increased more 40% compared to that before the lockdown. It was previously mentioned that the differences of the key meteorological parameters between the lockdown period and pre-lock period were marginal. Thus, it was supposed that the absolute contribution of natural surface emissions to GEM shouldn't be varying greatly between these two periods. However, even the mean temperature during the lockdown (6.1℃) was slightly lower than that before the lockdown (7.2℃), which was opposite to the PMF results on natural emission of GEM. Thus, we concluded that part of the increased contribution to GEM from natural release should be induced by the significantly reduced anthropogenic sources based on the analysis from different perspectives in this study.

16. Section 3.4: The uncertainty of the PMF results should be evaluated. For example, factor rotation could lead to very different outcome.

Response: Thanks for the suggestion. We have added the uncertainty of PMF results as shown in the Table below. By using the rotation tools in PMF, the Fpeak model run at the strength of 0.5, -0.5, 1, and -1 was conducted, respectively. We found that the changes in the Q value (dQ) due to the Fpeak rotation were less than 1% of the base run Q (robust) value. According to the user guide of PMF 5.0, it was acceptable when the percent for dQ was less than 5%. The profiles and contributions of each source were examined, and there were no significant differences between the factor contributions of base run and rotation results, especially for coal combustion and natural surface emission. Hence, the base run results were used in this study. In order to avoid confusion caused by the uncertainty of PMF results as much as possible, we take all resolved anthropogenic sources as a whole when compared with natural surface emission in the discussion.

We add sentences "To evaluate the uncertainty of the PMF results, the Fpeak model run at the strength of 0.5, -0.5, 1, and -1 were conducted by using the rotation tools in PMF. The changes of Q value (dQ) due to the Fpeak rotation were less than 1% of the base run Q (robust) value (Table S3), less than the benchmark value of 5%. The profiles and contributions of each source were also examined, and there were no significant differences between the factor contributions of base run and rotation results, especially for coal combustion and natural surface emission. Hence, the base run results were used in this study." in **Line 425-431** of the revision.

Table. Summary of Fpeak rotation and comparison of the source profiles and contribution between Base Run and Fpeak Run.

Fpeak strength	dQ (Robust)	%dQ (Robust)	natural surface emissions (%)		ship emission (%)		cement production (%)		iron and steel production (%)		vehicle emission (%)		coal combustion (%)	
			Base run	Fpeak run	Base run	Fpeak run	Base run	Fpeak run	Base run	Fpeak run	Base run	Fpeak run	Base run	Fpeak run
0.5	17	0.12	28.5	27.4	5.1	4.7	2.7	2.3	2.5	2.8	6.5	7.8	54.6	55.0
-0.5	14.8	0.11	28.5	28.3	5.1	5.1	2.7	4.0	2.5	1.9	6.5	6.4	54.6	54.2
1	65	0.47	28.5	26.8	5.1	4.6	2.7	1.6	2.5	3.1	6.5	7.7	54.6	56.1
-1	55.8	0.4	28.5	28.0	5.1	4.6	2.7	5.2	2.5	1.9	6.5	6.4	54.6	53.9

References:

Briggs, N. L. and Long, C. M.: Critical review of black carbon and elemental carbon source apportionment in Europe and the United States, Atmospheric Environment, 144, 409-427, 10.1016/j.atmosenv.2016.09.002, 2016.

Cheng, I., Xu, X., and Zhang, L.: Overview of receptor-based source apportionment studies for speciated atmospheric mercury, Atmos. Chem. Phys., 15, 7877-7895, 10.5194/acp-15-7877-2015, 2015.

Chong, X., Wang, Y., Liu, R., Zhang, Y., Zhang, Y., and Zheng, W.: Pollution characteristics and source difference of gaseous elemental mercury between haze and non-haze days in winter, The Science of the total environment, 678, 671-680, 10.1016/j.scitotenv.2019.04.338, 2019.

Fu, X. W., Zhang, H., Lin, C. J., Feng, X. B., Zhou, L. X., and Fang, S. X.: Correlation slopes of GEM / CO, GEM / CO₂, and GEM / CH₄ and estimated mercury emissions in China, South Asia, the Indochinese Peninsula, and Central Asia derived from observations in northwestern and southwestern China, Atmos. Chem. Phys., 15, 1013-1028, 10.5194/acp-15-1013-2015, 2015.

Huang, X., Ding, A., Gao, J., Zheng, B., Zhou, D., Qi, X., Tang, R., Wang, J., Ren, C., Nie, W., Chi, X., Xu, Z., Chen, L., Li, Y., Che, F., Pang, N., Wang, H., Tong, D., Qin, W., Cheng, W., Liu, W., Fu, Q., Liu, B., Chai, F., Davis, S. J., Zhang, Q., and He, K.: Enhanced secondary pollution offset reduction of primary emissions during COVID-19 lockdown in China, Natl Sci Rev, 8, nwaa137, 10.1093/nsr/nwaa137, 2021.

Lin, C. J., Zhu, W., Li, X. C., Feng, X. B., Sommar, J., and Shang, L. H.: Novel Dynamic Flux Chamber for Measuring Air-Surface Exchange of Hg-o from Soils, Environmental science & technology, 46, 8910-8920, 10.1021/es3012386, 2012.

Liu, K., Wu, Q., Wang, L., Wang, S., Liu, T., Ding, D., Tang, Y., Li, G., Tian, H., Duan, L., Wang, X., Fu, X., Feng, X., and Hao, J.: Measure-Specific Effectiveness of Air Pollution Control on China's Atmospheric Mercury Concentration and Deposition during 2013-2017, Environmental science & technology, 10.1021/acs.est.9b02428, 2019.

Pannu, R., Siciliano, S. D., and O'Driscoll, N. J.: Quantifying the effects of soil temperature, moisture and sterilization on elemental mercury formation in boreal soils, Environmental Pollution, 193, 138-146, 10.1016/j.envpol.2014.06.023, 2014.

Streets, D. G., Horowitz, H. M., Lu, Z., Levin, L., Thackray, C. P., and Sunderland, E. M.: Global and regional trends in mercury emissions and concentrations, 2010–2015, Atmospheric Environment, 201, 417-427, 10.1016/j.atmosenv.2018.12.031, 2019.

Wang, X., Lin, C.-J., Yuan, W., Sommar, J., Zhu, W., and Feng, X.: Emission-dominated gas exchange of elemental mercury vapor over natural

surfaces in China, Atmospheric Chemistry and Physics, 16, 11125-11143, 10.5194/acp-16-11125-2016, 2016.

Wu, Q., Li, G., Wang, S., Liu, K., and Hao, J.: Mitigation Options of Atmospheric Hg Emissions in China, Environmental science & technology, 52, 12368-12375, 10.1021/acs.est.8b03702, 2018.

Wu, Q., Tang, Y., Wang, L., Wang, S., Han, D., Ouyang, D., Jiang, Y., Xu, P., Xue, Z., and Hu, J.: Impact of emission reductions and meteorology changes on atmospheric mercury

concentrations during the COVID-19 lockdown, The Science of the total environment, 750, 142323, 10.1016/j.scitotenv.2020.142323, 2021.
Xin, M. and Gustin, M. S.: Gaseous elemental mercury exchange with low mercury containing soils: Investigation of controlling factors, Appl. Geochem., 22, 1451-1466, 10.1016/j.apgeochem.2007.02.006, 2007.
Yang, J., Wen, Y., Wang, Y., Zhang, S., Pinto, J. P., Pennington, E. A., Wang, Z., Wu, Y., Sander, S. P., Jiang, J. H., Hao, J., Yung, Y. L., and Seinfeld, J. H.: From COVID-19 to future electrification: Assessing traffic impacts on air quality by a machine-learning model, Proc Natl Acad Sci U S A, 118, 10.1073/pnas.2102705118, 2021.
Yang, Y. and Zhao, Y.: Quantification and evaluation of atmospheric pollutant emissions from open biomass burning with multiple methods: a case study for the Yangtze River Delta region, China, Atmos. Chem. Phys., 19, 327-348, 10.5194/acp-19-327-2019, 2019.
Zhang, L., Wang, S. X., Wang, L., and Hao, J. M.: Atmospheric mercury concentration and chemical speciation at a rural site in Beijing, China: implications of mercury emission sources, Atmospheric Chemistry and Physics, 13, 10505-10516, 10.5194/acp-13-10505-2013, 2013.
Zhang, L., Zhou, P. S., Zhong, H., Zhao, Y., Dai, L., Wang, Q. G., Xi, M. X., Lu, Y., and Wang, Y. T.: Quantifying the impacts of anthropogenic and natural perturbations on gaseous elemental mercury (GEM) at a suburban site in eastern China using generalized additive models, Atmospheric Environment, 247, 10.1016/j.atmosenv.2020.118181, 2021.
Zhang, Y., Jacob, D. J., Horowitz, H. M., Chen, L., Amos, H. M., Krabbenhoft, D. P., Slemr, F., St Louis, V. L., and Sunderland, E. M.: Observed decrease in atmospheric mercury explained by global decline in anthropogenic emissions, Proc Natl Acad Sci U S A, 113, 526-531, 10.1073/pnas.1516312113, 2016.
Zhu, J., Wang, T., Talbot, R., Mao, H., Hall, C. B., Yang, X., Fu, C., Zhuang, B., Li, S., Han, Y., and Huang, X.: Characteristics of atmospheric Total Gaseous Mercury (TGM) observed in urban Nanjing, China, Atmospheric Chemistry and Physics, 12, 12103-12118, 10.5194/acp-12-12103-2012, 2012.
Zhu, W., Lin, C. J., Wang, X., Sommar, J., Fu, X., and Feng, X.: Global observations and modeling of atmosphere–surface exchange of elemental mercury: a critical review, Atmos. Chem. Phys., 16, 4451-4480, 10.5194/acp-16-4451-2016, 2016.

Response to Reviewer #2's Comments

Anonymous Referee #2:

Major comments:

This work investigates how anthropogenic and natural mercury emissions differ before, during, and after the COVID-19 lockdown. The paper aims to show that the decrease in anthropogenic mercury emissions during the lockdown led to an increase in natural release. The methods used to quantify that are correlation analysis, a PMF model and a neural network (NN) combined with SHAP values. The NN learned to predict gaseous elemental mercury (GEM) given other air pollutants and atmospheric conditions and then applies the SHAP approach to obtain how much each input feature contributed to the prediction. The choice of method, using an NN with SHAP values, seems suitable for this setting. The performance of the NN is not very good, this will also affect the interpretability of the SHAP values. In general, more details of the NN would be helpful.

From my point of view, this paper needs rewriting and changes before it can be accepted for publication at ACP.

We sincerely thank for the reviewer's in-depth comments and helpful suggestions on this manuscript. Based on the specific comments, we have responded to all the comments point-by-point and made corresponding changes in the manuscript as highlighted in red color. The reviewer has raised a number of issues and we quite agree. We feel the substantial revisions based on the reviewer's comments have greatly improved the quality of this manuscript. Please check the detailed responses to all the comments as below.

Major comments:

(1). The performance of the NN is not very good, this will also affect the interpretability of the SHAP values.

Response: Thanks for the comment. Compared to numerous machine learning studies applied in common air pollutants such as $PM_{2.5}$, the performance of NN in simulating mercury is indeed not so good. There are some reasons. First, for those common air pollutants, the availability of the monitoring network (e.g., spatial coverage and time span of measurement) is sufficient. Thus, there are enough data for the training. As a comparison, there is only one multi-year mercury observational site in the Yangtze River Delta region. On the other hand, the model performance depends a lot on the input parameters. By taking building the $PM_{2.5}$ model as an example, parameters such as aerosol optical depth from remote sensing products are usually considered. As aerosol optical depth is directly related to $PM_{2.5}$, the model performance usually can be improved to a certain extent. In contrast, there is no such parameter for mercury and this may limit the goodness of model performance.

In this study, as shown in the figure below, the R^2 value between ANN-simulated and observed hourly GEM concentration was 0.67. By averaging the hourly data as the daily data, the R^2 value increased to 0.85, and the modeled GEM concentration was ~10% underestimated compared to observation. This performance is better than previous study using CMAQ model in eastern China, which found that the modeled TGM concentrations in Shanghai was ~51% overestimated (Zhu et al., 2015). In addition, to test the applicability of the model on the regional scale, we compared the observed and simulated GEM concentrations in Suzhou, Ningbo, Nanjing, and Hefei (Figure below). In Nanjing and Suzhou, the observed and simulated daily GEM showed consistence with R^2 values of 0.52 and 0.71, respectively. In Ningbo, the observed and simulated GEM in summer and winter also showed consistence with R^2 values of 0.64 and 0.65, respectively. A low bias was derived between the observed and simulated seasonal GEM in Hefei. This suggested that it is feasible to use the trained ANN model to simulated the GEM concentrations in Shanghai and even the Yangtze River Delta region.

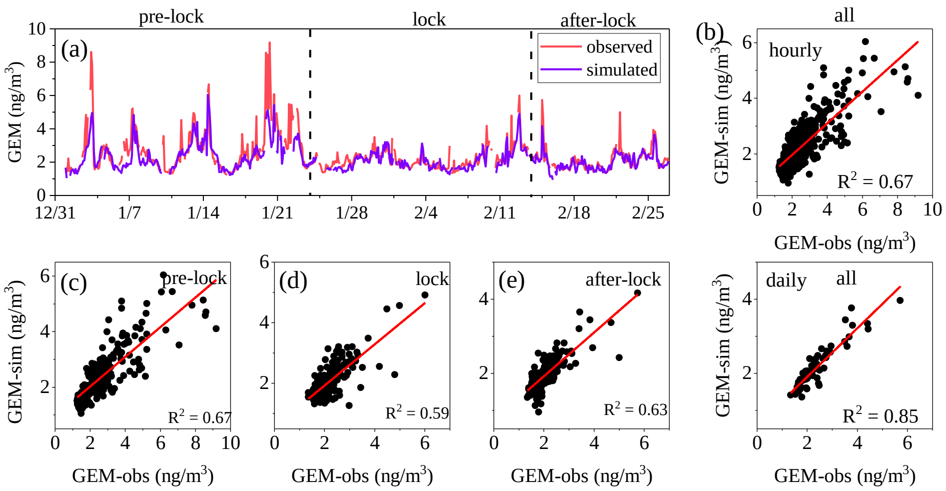

Figure. (a) Time-series of observed and ANN-simulated GEM concentrations during the study period. (b) Linear correlation between observed and ANN-simulated GEM concentrations. (c)-(e) Linear correlation between observed and ANN-simulated GEM concentrations before, during, and after the lockdown.

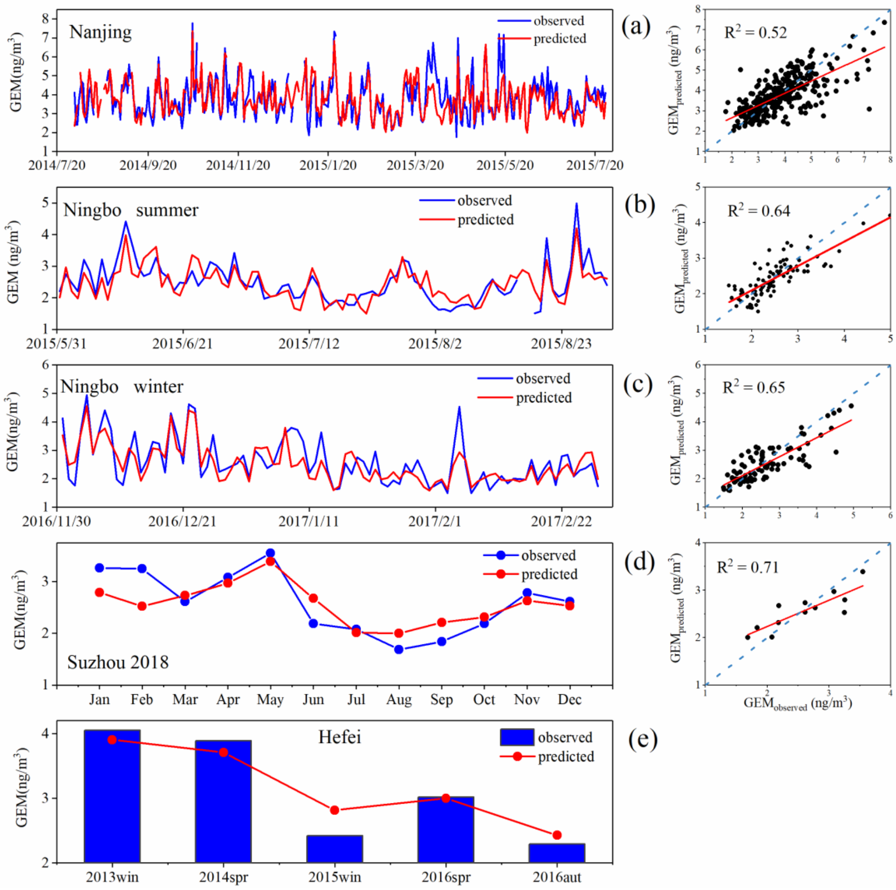

Figure. Comparison of observed (blue line) and simulated (red line) GEM concentrations in (a) Nanjing, (b) summer Ningbo, (c) winter Ningbo, (d) Suzhou, and (e) Hefei. The left panel is their time series, and the right panel is their corresponding correlation. The blue dotted line represents the y = x reference line.

In the revised manuscript (**Line 494-500**), we have added a paragraph about the shortcomings of this study and some future outlooks in the conclusion section.

"However, we realize that the performance of machine learning in simulating atmospheric mercury in this study has yet to be improved. Continuous long-term observations of atmospheric mercury with more monitoring sites are desired to ensure a more adequate training dataset. Also, more relevant environment parameters for GEM are needed to further improve the training performance of machine learning model. In addition, different machine learning methods such as artificial neural network, decision tree, random forest, and Bayesian learning should be evaluated to choose an optimal solution.".

(2). In general, more details of the NN would be helpful.

Response: Thanks for the suggestion. We have added the detailed description of the training and validation of the ANN model in **Line 175-193** of the revised manuscript.

"The long-term observational GEM (hourly data from March 1, 2015 to February 28, 2019; n = 17532) in Shanghai was the target variable for training, and the corresponding air pollutants (SO_2, CO, O_3, NO_2, and $PM_{2.5}$) and meteorological data (air temperature, relative humidity, and wind speed) were chosen as input variables for training. The datasets were randomly divided into three parts, i.e., 70% for training, 15% for validation, and 15% for testing. We chose the neural network containing a hidden layer with 20 nodes, and the training algorithm was the Levenberg-Marquardt. The performance of the model was evaluated with the mean square error (MSE) and correlation coefficient (R^2 value). To verify the accuracy of the trained neural network model, we compared the observed (not included in the training data set) and simulated GEM concentrations of DSL from January 1 to February 26, 2020, and found that they exhibited a reasonably good correlation with the correlation coefficient (R^2) of 0.67. To test the applicability of the model on the regional scale, we compared the observed and simulated GEM concentrations in Suzhou, Ningbo, Nanjing, and Hefei (Figure S2). In Nanjing and Suzhou, the observed and simulated daily GEM showed consistence with R^2 values of 0.52 and 0.71, respectively. In Ningbo, the observed and simulated GEM in summer and winter also showed consistence with R^2 values of 0.64 and 0.65, respectively. A low bias was derived between the observed and simulated seasonal GEM in Hefei. This suggested that it was feasible to use the trained ANN model to simulated the GEM concentrations in Shanghai and even the Yangtze River Delta region."

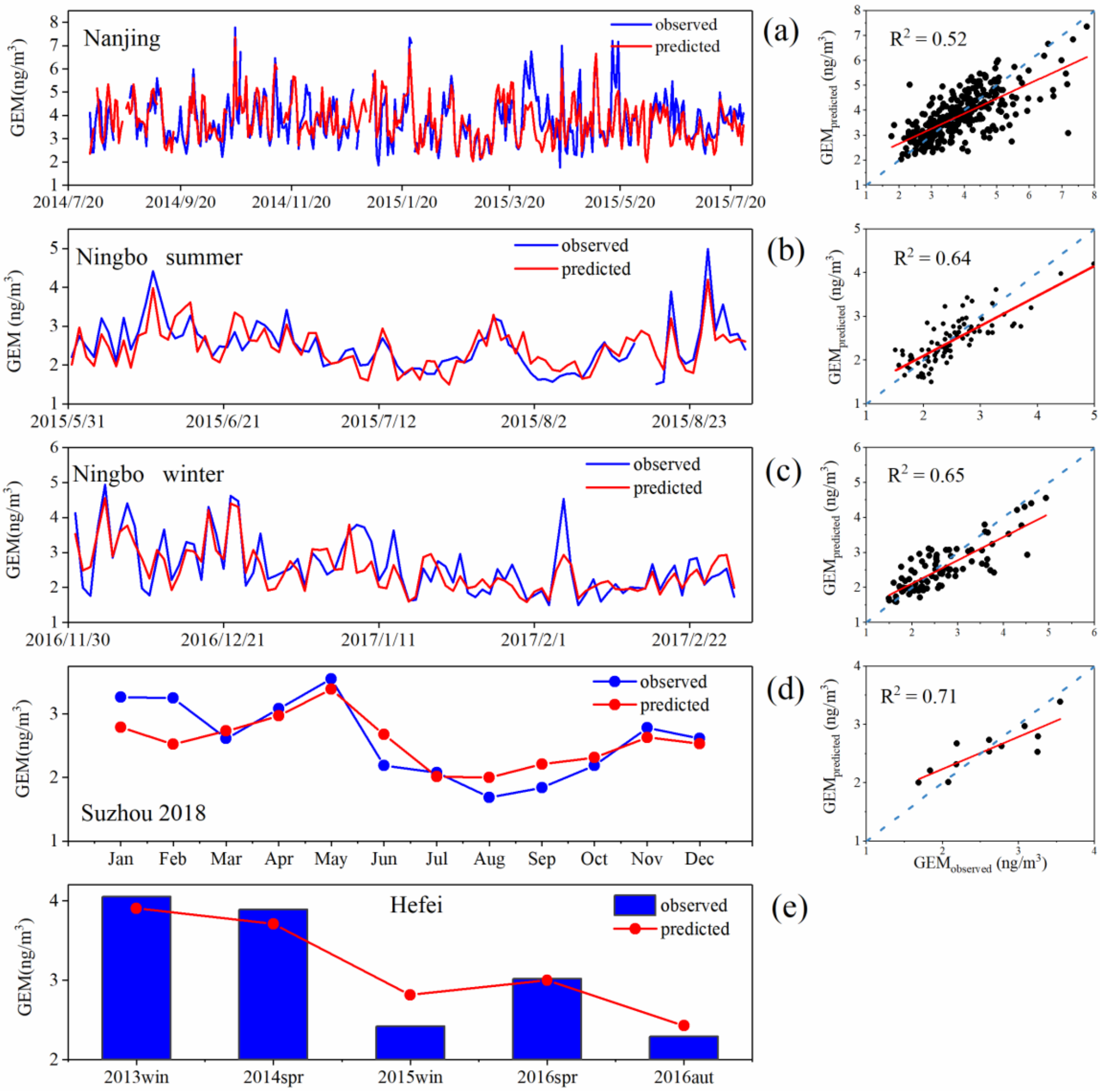

Figure S2. Comparison of observed (blue line) and simulated (red line) GEM concentrations in (a) Nanjing, (b) summer Ningbo, (c) winter Ningbo, (d) Suzhou, and (e) Hefei. The left panel is their time series, and the right panel is their corresponding correlation. The blue dotted line represents the y = x reference line.

Specific comments:

1. Line 106: The PMF approach should be mentioned in the introduction

Response: Thanks for the suggestion. We have added descriptions about the PMF approach in the introduction section in **Line 101-116** of the revised manuscript.

"Many receptor – based models have been used to determine the sources and processes of air pollutants, among which the positive matrix factorization (PMF) is a commonly used method (Yu et al., 2019; Sun et al., 2016; Chang et al., 2018). The PMF method provides quantitative source profiles and source contributions, and the obtained source profiles can aid factor interpretation (Belis et al., 2013). Another strength of PMF is that the measurement uncertainty is included in the PMF model, which ensures that species with large uncertainties have less impact on the model results (Hopke, 2016). Many previous studies have applied the PMF method to the source apportionment of atmospheric mercury. One study in Canada compared the PMF model performance of atmospheric mercury in different years and found that the source profiles and source contributions of GEM in 2009 and 2010 were in good agreement (Xu et al., 2017). By using the PMF model, the research on the western coast of Ireland found that baseline and combustion processes were the controlling sources of atmospheric mercury (Custodio et al., 2020). The study in the Yangtze River Delta in eastern China suggested that the contribution of natural sources to GEM had gradually exceeded that of anthropogenic sources from 2015 to 2018 by using the PMF method (Qin et al., 2020). This indicated that it is feasible to use the PMF model to identify the source of GEM in the atmosphere."

2. Line 155: Please include a few details of the NN in this work, such as data size, how the train-val-test split was done, and a comment on hyperparameter tuning

Response: Thanks for the suggestion. We have added the detailed description of the training and validation of the ANN model in **Line 175-193** of the revised manuscript.

"The long-term observational GEM (hourly data from March 1, 2015 to February 28, 2019; n = 17532) in Shanghai was the target variable for training, and the corresponding air pollutants (SO_2, CO, O_3, NO_2, and $PM_{2.5}$) and meteorological data (air temperature, relative humidity, and wind speed) were chosen as input variables for training. The datasets were randomly divided into three parts, i.e., 70% for training, 15% for validation, and 15% for testing. We chose the neural network containing a hidden layer with 20 nodes, and the training algorithm was the Levenberg-Marquardt. The performance of the model was evaluated with the mean square error (MSE) and correlation coefficient (R^2 value). To verify the accuracy of the trained neural network model, we compared the observed (not included in the training data set) and simulated GEM concentrations of DSL from January 1 to February 26, 2020, and found that they exhibited a reasonably good correlation with the correlation coefficient (R^2) of 0.67. To test the applicability of the model on the regional scale, we compared the observed and simulated GEM concentrations in Suzhou, Ningbo, Nanjing, and Hefei (Figure S2). In Nanjing and Suzhou, the observed and simulated daily GEM showed consistence with R^2 values of 0.52 and 0.71, respectively. In Ningbo, the observed and simulated GEM in summer and winter also showed consistence with R^2 values of 0.64 and 0.65, respectively. A low bias was derived between the observed and simulated seasonal GEM in Hefei. This suggested that it was feasible to use the trained ANN model to simulated the GEM concentrations in Shanghai and even the Yangtze River Delta region."

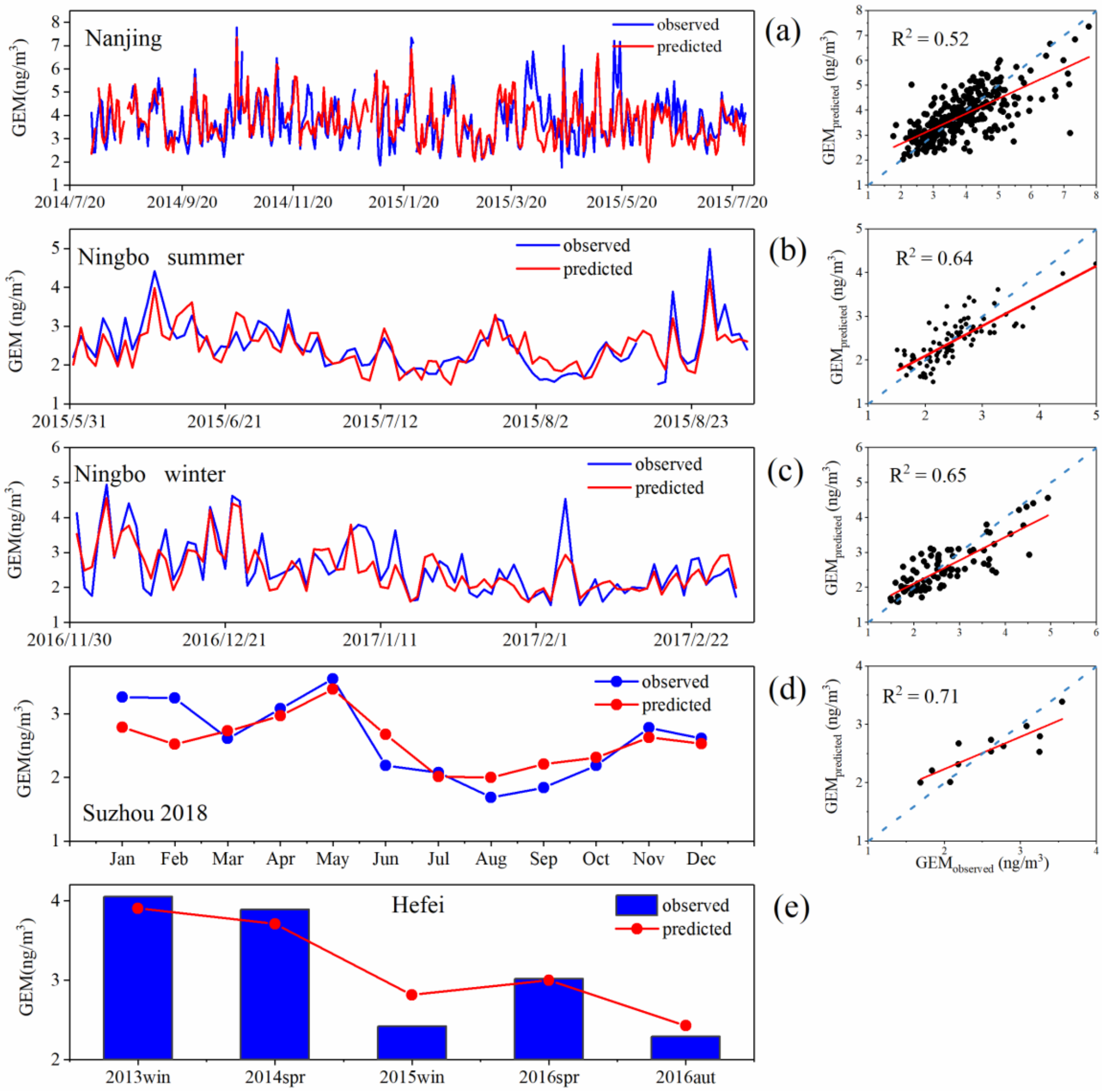

Figure S2. Comparison of observed (blue line) and simulated (red line) GEM concentrations in (a) Nanjing, (b) summer Ningbo, (c) winter Ningbo, (d) Suzhou, and (e) Hefei. The left panel is their time series, and the right panel is their corresponding correlation. The blue dotted line represents the y = x reference line.

3. Line 284: R^2 values are missing in three out of the nine subplots

Response: Thanks for pointing this out. We have added the missing R^2 value to the subplots as shown in the figure below.

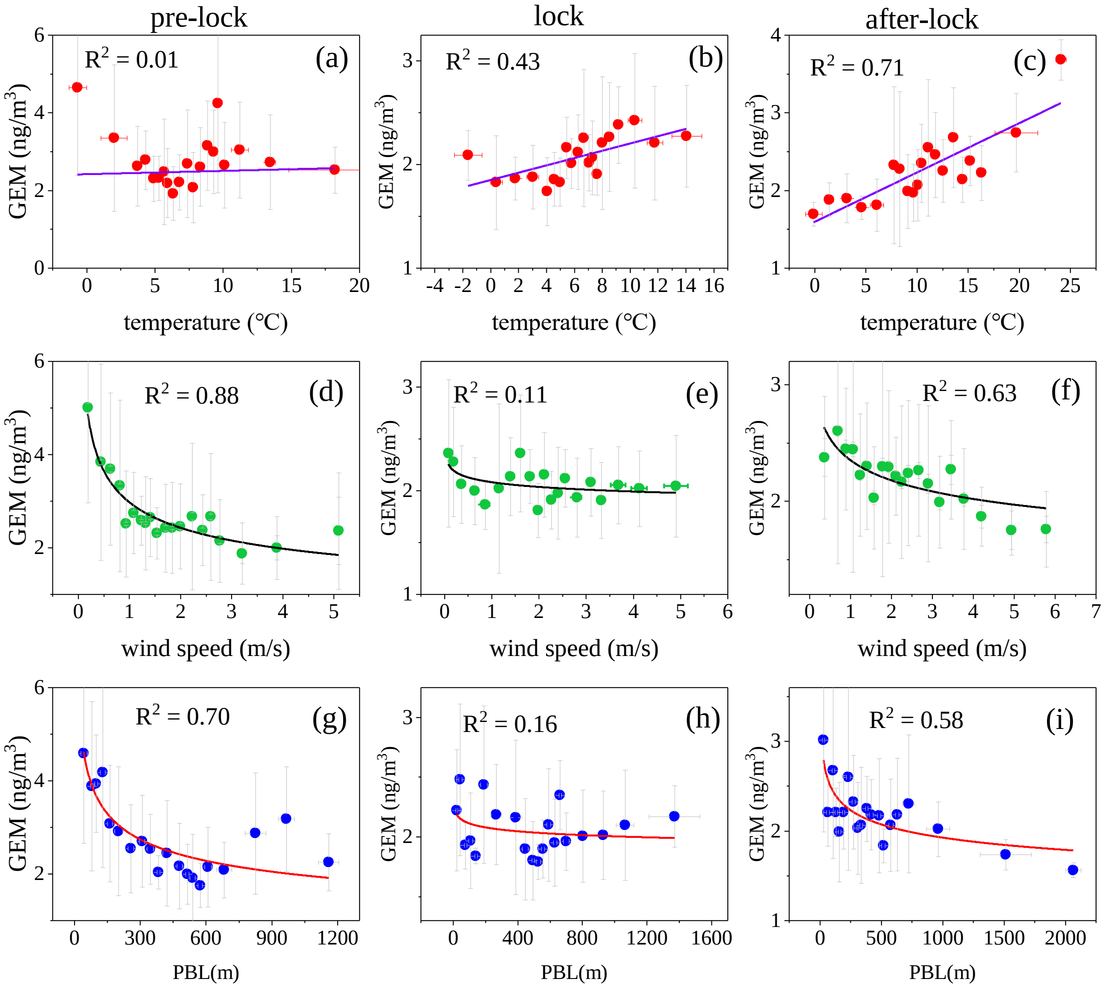

Figure. Relationship between GEM concentration and (a-c) temperature, (d-f) wind speed, and (g-i) PBL height before, during, and after the lockdown.

4. Line 284: What are the lines shown in the plots and why aren't they shown in each of the subplots

Response: The lines in the plots represent the fitting lines of the scatter points. Some subplots are not shown with fitting lines as the investigated parameters are poorly correlated. In the revision, we have added the missing fitting lines and R^2 value to the subplots as shown in the figure below.

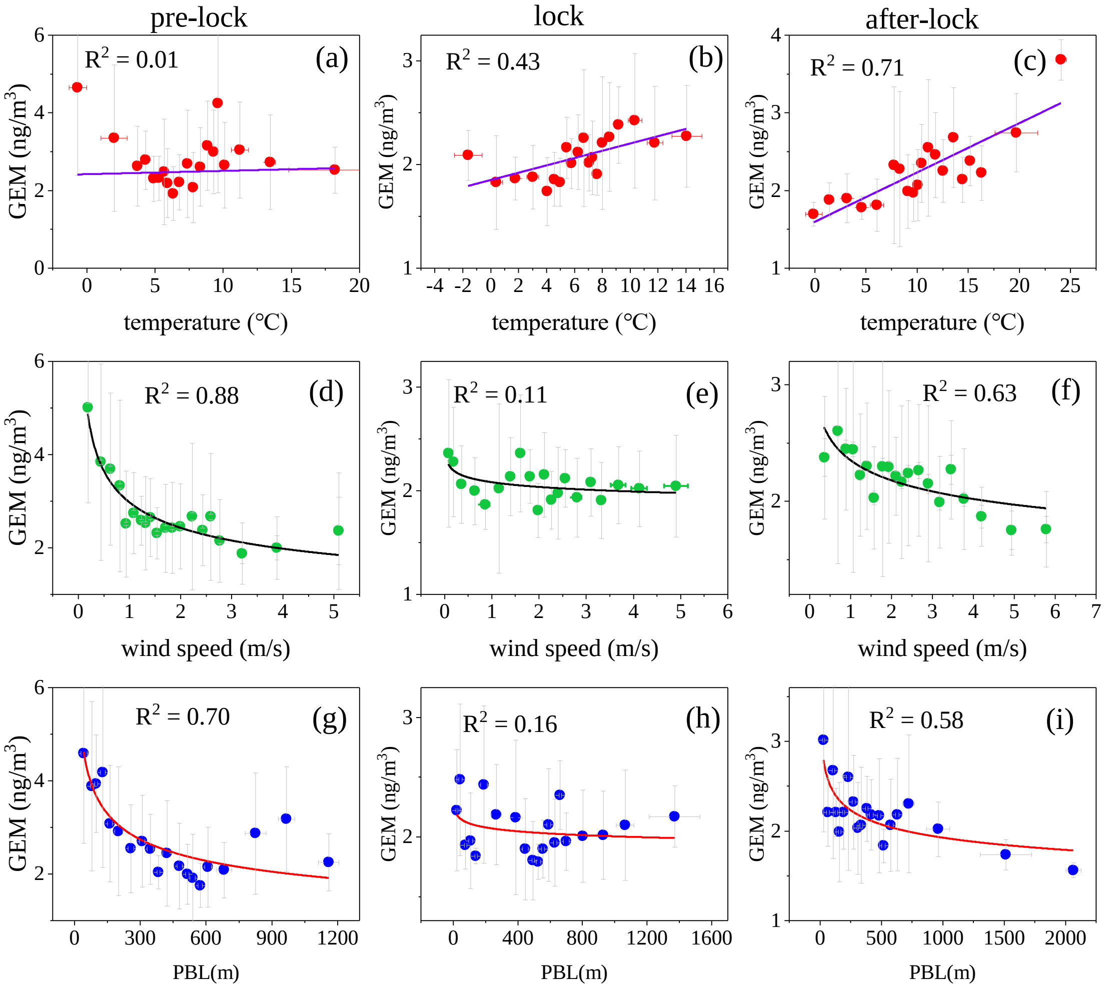

Figure. Relationship between GEM concentration and (a-c) temperature, (d-f) wind speed, and (g-i) PBL height before, during, and after the lockdown.

5. Line 293: The R^2 score of 0.67 is okay, but not great.

Response: This sentence is changed as “Figure 4a-b shows the comparison of ANN-simulated and observed GEM concentrations during the whole study period, and found their correlation coefficient is acceptable (R^2 = 0.67).” in **Line 351-353** of the revised manuscript.

6. Line 293: It would be relevant to know the different performances of the NN evaluated on the pre-lockdown, lockdown, and post-lockdown periods.

Response: Thanks for the suggestion. We examined the performance of the ANN model before, during, and after the lockdown. As shown in the figure below, the correlation between ANN simulated and observed GEM concentrations were 0.67, 0.59, and 0.63, respectively.

We added the sentence “As shown in Figure S5, we examined the performance of the ANN model before, during, and after the lockdown. The correlations between ANN simulated and observed GEM concentrations were also acceptable with correlation coefficient of 0.67, 0.59, and 0.63, respectively.” in **Line 353-356** of the revision.

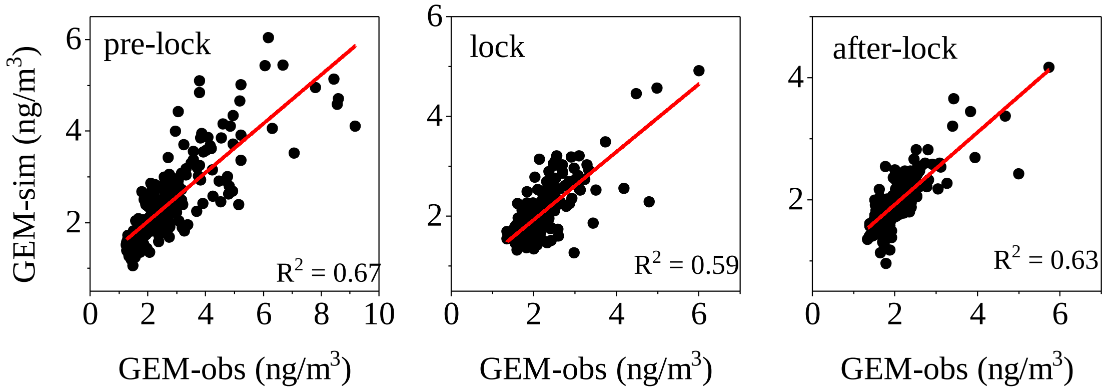

Figure S5. Linear correlation between observed and ANN-simulated GEM concentrations before, during, and after the lockdown.

References:

Belis, C. A., Karagulian, F., Larsen, B. R., and Hopke, P. K.: Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe, Atmospheric Environment, 69, 94-108, 10.1016/j.atmosenv.2012.11.009, 2013.

Chang, Y. H., Huang, K., Xie, M. J., Deng, C. R., Zou, Z., Liu, S. D., and Zhang, Y. L.: First long-term and near real-time measurement of trace elements in China's urban atmosphere: temporal variability, source apportionment and precipitation effect, Atmospheric Chemistry and Physics, 18, 11793-11812, 10.5194/acp-18-11793-2018, 2018.

Custodio, D., Ebinghaus, R., Spain, T. G., and Bieser, J.: Source apportionment of atmospheric mercury in the remote marine atmosphere: Mace Head GAW station, Irish western coast, Atmospheric Chemistry and Physics, 20, 7929-7939, 10.5194/acp-20-7929-2020, 2020.

Hopke, P. K.: Review of receptor modeling methods for source apportionment, Journal of the Air & Waste Management Association, 66, 237-259, 10.1080/10962247.2016.1140693, 2016.

Qin, X., Zhang, L., Wang, G., Wang, X., Fu, Q., Xu, J., Li, H., Chen, J., Zhao, Q., Lin, Y., Huo, J., Wang, F., Huang, K., and Deng, C.: Assessing contributions of natural surface and anthropogenic emissions to atmospheric mercury in a fast-developing region of eastern China from 2015 to 2018, Atmos. Chem. Phys., 20, 10985-10996, 10.5194/acp-20-10985-2020, 2020.

Sun, Y., Chen, C., Zhang, Y., Xu, W., Zhou, L., Cheng, X., Zheng, H., Ji, D., Li, J., Tang, X., Fu, P., and Wang, Z.: Rapid formation and evolution of an extreme haze episode in Northern China during winter 2015, Scientific reports, 6, 27151, 10.1038/srep27151, 2016.

Xu, X., Liao, Y., Cheng, I., and Zhang, L.: Potential sources and processes affecting speciated atmospheric mercury at Kejimkujik National Park, Canada: comparison of receptor models and data treatment methods, Atmospheric Chemistry and Physics, 17, 1381-1400, 10.5194/acp-17-1381-2017, 2017.

Yu, Y., He, S., Wu, X., Zhang, C., Yao, Y., Liao, H., Wang, Q., and Xie, M.: PM2.5 elements at an urban site in Yangtze River Delta, China: High time-resolved measurement and the application in source apportionment, Environ Pollut, 253, 1089-1099, 10.1016/j.envpol.2019.07.096, 2019.

Zhu, J., Wang, T., Bieser, J., and Matthias, V.: Source attribution and process analysis for atmospheric mercury in eastern China simulated by CMAQ-Hg, Atmos. Chem. Phys., 15, 8767-8779, 10.5194/acp-15-8767-2015, 2015.